TEXT FLY WITHIN
THE BOOK ONLY

UNIVERSAL
LIBRARY

OU_156746

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 530/D22N Accession No. 14825

Author Darwin, C. G.

Title New conceptions of matter

This book should be returned on or before the date last marked below.

THE
NEW CONCEPTIONS OF MATTER

BY

C. G. DARWIN, M.A., F.R.S.
TAIT PROFESSOR OF NATURAL PHILOSOPHY
IN THE UNIVERSITY OF EDINBURGH

LONDON
G. BELL AND SONS, LTD.
1931

DEDICATED AS A TRIBUTE OF
ADMIRATION AND AFFECTION

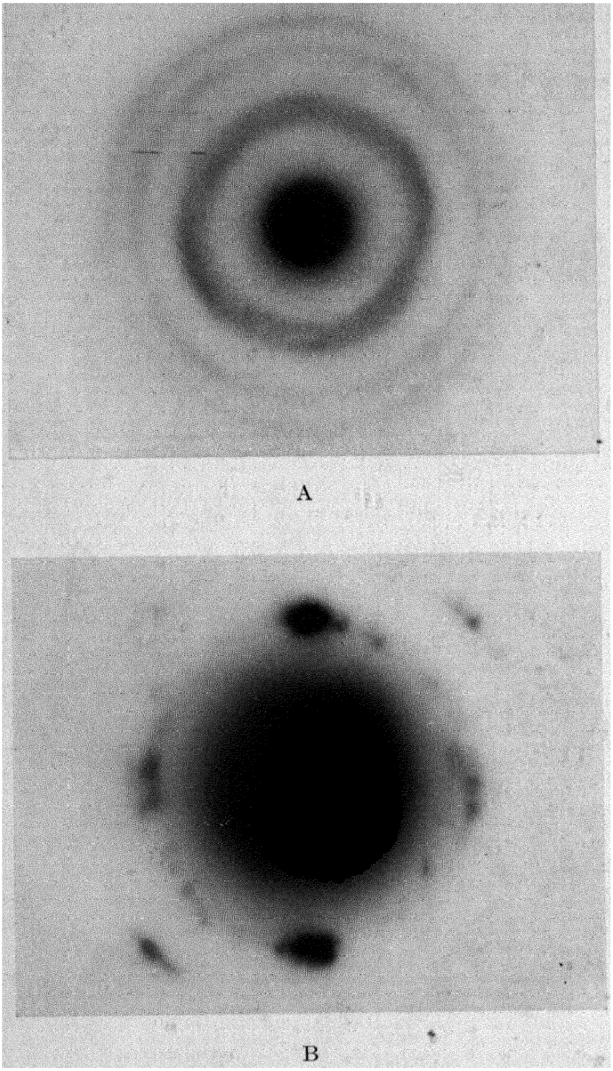
TO

NEILS BOHR

THE ACKNOWLEDGED LEADER
AMONG THOSE WHO HAVE SHOWN
HOW THE SECRETS OF NATURAL PHILOSOPHY
MAY BE UNDERSTOOD

Printed in Great Britain by
Neill & Co., Ltd., Edinburgh.

THE NEW CONCEPTIONS
OF MATTER



Diffraction of Electrons

A beam of electrons passes through a thin film of metal, and then strikes a photographic plate. The rings are due to the diffraction of the electrons by the small crystals of the metal. (*Photographs by Prof G P. Thomson*)

A. Gold.

B. Aluminium. Some of the crystals are rather large and consequently produce more intense spots on the rings.

PREFACE

In the spring of the present year I delivered a course of lectures at the Lowell Institute of Boston, and the text of the present work was prepared in connection with those lectures. In lecturing it is of course necessary to divide the subject into roughly equal lengths, and I have adhered to this arrangement, even though in a more systematic development whole sections would perhaps have been transferred from one chapter to another. I do not know how far my presentation will find favour with experts; it is certainly not the same as that adopted in some of the more technical works that have recently appeared, and yet I believe that it makes a consistent formulation of the foundations, which only needs the help of mathematics to yield all the results of the theory so far obtained. However that may be, I have little hesitation in saying that, if it is to be at all possible to present the new mechanics in popular or even semi-popular terms, it must be done more or less on the lines I have adopted here.

Some of the illustrations have been provided by friends, and I wish to thank Professors G. P. Thomson and E. N. da C. Andrade and Mr P. Blackett for their kind help in the matter. I wish also to thank my wife for many helpful criticisms in the construction of the work.

C. G. D.

August 1931.

CONTENTS

CHAP. PAGE

INTRODUCTION 1

I. THE STARTING-POINT 8

Atoms – Electrons – The Nucleus – Protons – Photons – The Æther – Momentum and Energy.

II. WAVES 29

Harmonic Waves – The Principle of Superposition – Diffraction – Rays of Light – Groups of Waves.

III. THE DIFFRACTION OF MATTER 53

The Principle of Least Action – The Propagation of Waves in Variable Media – Diffraction Gratings – X-rays and Crystals – Diffraction of Electrons.

IV. THE UNCERTAINTY PRINCIPLE 78

Contrast of Wave and Particle – Wave-length and Momentum – Uncertainty of Speed and Position – Experiment with Microscope – Pressure of Light – Other Experiments – Causality.

V. THE ATOM 103

Types of Wave – Electron near Nucleus – Normal Modes of Vibration – Modes of Hydrogen Atom – Relation of Modes to Orbits – Energy Levels.

VI. POLARISATION 125

Character of Polarised Light – Doubling of Energy Levels – Spinning Electron – Stern-Gerlach Experiment – Polarisation of Free Electron – Polarisation of Proton.

CHAP. PAGE

VII. COLLISIONS 146

Wave of Two Particles – Scattering of Electrons by Fixed Nucleus – Collision of Two Free Particles – Various Collisions of α-particles – Compton Effect – Optical Effects.

VIII. THE EXCLUSION PRINCIPLE 168

Interaction of Identical Systems – Symmetry and Antisymmetry – Antisymmetry of Electrons and Protons – Collision of Two α-particles – Periodic Table of Elements – Magnetism of Iron – Conduction of Electricity.

INDEX 191

LIST OF PLATES

The Diffraction of Electrons . . *Frontispiece*

I. Alpha-ray Tracks . . . . *Facing page* 15

II. Young's Experiment . . . „ „ 40

III. A. The Diffraction of X-rays . . „ „ 74

B. The Diffraction of Electrons .

IV. The Stern-Gerlach Experiment . „ „ 140

V. Alpha-ray Collisions . . . „ „ 157

THE NEW CONCEPTIONS OF MATTER

INTRODUCTION

It is one of the most unsatisfactory features of the enormous recent developments of science, that they are so remote from all the ordinary things of life. In a few branches of science it is possible to explain to the layman what one is doing in a few words—everyone can appreciate the work of the medical researcher who is trying to cure the common cold, or of the chemist who is studying how to keep steel from rusting. But there is very little work in science that is so simple as that; for probably the man who is going to cure our colds will have spent twenty years in finding out how to keep a certain microscopic organism alive, and the chemist may be studying the effect of making the walls of a blast furnace of some different material. Neither of these occupations would seem to the ordinary man to have any special value, and the scientist often finds it hard to explain the true situation, which is that he does his work without ulterior motive because he finds it interesting. Moreover he is a modest man, who knows that very probably nothing will come of his work, and so he does not explain that if it succeeds it will help some other work which may help some further work, which might then turn out to be of practical value.

But the position may become a dangerous one on both sides, for the layman may come to look on the scientist as the priest of an esoteric religion which he respects but does not understand, or perhaps even want to understand, and again the scientist is sometimes also a little inclined to claim the attributes of the priest, and regard himself as the master of mysteries which will for ever be beyond the ken of the flock. Apart from the fact that this makes him a very insufferable person, it has the dangers of isolation. There are many cases in history of scientific schools which have worked with extraordinary energy and intelligence on subjects which seem entirely trivial to us now. We must recognise that almost any subject in the world, if studied deeply enough, becomes interesting, and that this exposes us all to the insidious dangers of pedantry, for the pedant is only the learned man whose mind is engaged with unimportant things. Therefore the scientist should be ready sometimes to justify the importance, I certainly do not wish to say the utility, of his studies before the general tribunal.

In this book it is my aim to try and clear myself of the imputation of esotericism by describing, as far as I may, the ideas at present current about the basic constituents of the physical world. There have been several excellent books recently written on the subject, but it has seemed to me that some of them, perhaps through attempting to cover too wide a field, do not always succeed in explaining their subject. They excite the wonder of the reader, by suggesting to him what extraordinary difficulties there are in the ideas of physics; they are like a conjuror whose tricks seem to us inexplicable. I have set myself what is, I think, a more ambitious task, for I want to try and show how the tricks are done. I shall count myself as having succeeded, if at the end of the book any surviving reader will speak no longer of the mysteries of science, but, shall we say, of the naturalness of Nature.

This is not meant as a depreciation of the gigantic efforts of intellect which have been needed in order to achieve the current theories; it would be a poor testimonial to humanity if it took nearly three centuries to solve a problem that was really easy. No, the difficulties were enormous, but they lay chiefly in guessing which of the countless phenomena of nature would supply the key, and how the facts could be freed from the mass of really irrelevant though apparently inseparable detail. But once some genius has found the way it ought to be possible to follow it, without the necessity of any deep technical knowledge in doing so.

A great part of the difficulty is attributable to the technique of discovery, which in physics has been usually mathematical. In ordinary life we act, or ought to act, on the principles of common sense, that is to say, on generalisation of our own previous experiences. The discoverer of a really new principle has nothing of the kind to guide him, but must fall back on logic, the only faculty on which we feel we can always place reliance. The mathematical form in which new theories are usually dressed is the method by which we assure ourselves that our reasoning is logically sound, and therefore that it can be applied in unfamiliar situations. In his admirable work, *The Nature of the Physical World*, Eddington has said that whereas in the nineteenth century the Creator was regarded as an engineer, in the twentieth he is a pure mathematician. I do not agree with this theological diagnosis. Every *new* body of discovery is mathematical in form, because there is no other guidance that we can have. But as familiarity grows we find unsuspected analogies with our previous experience, and the engineer gets his chance of replacing the mathematician. When Eddington spoke of the nineteenth century he was thinking of the middle and end of it; if he had looked at the beginning he would have seen that the great discoveries of optical theory began life in mathematical form too. Now the new discoveries which I am to describe were only begun five years ago, and so it is hardly surprising that they are still partly covered with the scaffolding of mathematical formalism. But I hope to show that the time has already come, when it is possible to free the structure from a good deal of this scaffolding and so to gain something of an intuitive view of what the world is really like.

In the understanding of the new theory, the difficulties cannot by any means be all attributed to the mathematicians. As in the history of most discoveries, the path has been very devious, and the physicists must bear the responsibility for many of its wanderings. The earlier stages of the development were largely devoted to the study of the character of the light emitted by gases. This is a very intricate matter, but it brought a great reward, in that it became possible through it to understand in a general way how the interior of an atom is put together, and this knowledge was the chief thing that led to the new principles. The historical course of development has not, unnaturally, had rather a strong influence on our outlook, and has perhaps given to some the impression that the central feature of the new mechanics is the construction of atoms, and therefore that the subject is bound to be a complicated one. Here I shall chiefly be concerned with general principles, and it will, I hope, be seen that this complicated approach is unnecessary.

The truth is that physics, the parent science, has once more given birth to a new science. Some fifty years ago, by the application of thermodynamics to chemical problems, physics brought forth the great science of physical chemistry. Twenty years ago, by the study of crystals with X-rays, the formal geometrical theory of crystallography was transformed, and the science of physical crystallography was born. Now again a new science has emerged from the union with inorganic chemistry of the new ideas of physics. We may perhaps make a discrimination between physics and chemistry by saying that physics is concerned with the resemblances between things, and chemistry with their differences, and I will therefore call the new science mathematical chemistry. It is not a perfect description, but the more appropriate name has already been given to its elder brother. It is understating the case to say that it is a healthy child, for it is at present attracting more attention than its parent. It is hardly an exaggeration that in any of the current physical journals three-quarters of the papers are concerned with mathematical chemistry. It is a subject which calls for the highest qualities of mathematical ability and physical insight, and the majority of physicists are still engaged on its fascinating problems; but I shall hardly touch on it, for the physical foundations on which it rests will occupy all our time. In describing those principles it will be necessary to enter in some detail into many things which the professional scientist may regard as rather old-fashioned. I do not apologise, therefore, for the fact that more than one of the chapters will be devoted to subjects which would have been described in much the same way fifty or more years ago; for with the new insight that has now been gained, these matters become the very kernel of our subject.

When we approach any new branch of knowledge we nearly always find that the things which interest the specialists are not the obvious things which we should expect. Thus we might expect the law to be chiefly concerned with such questions as the penalty for murder, whereas in fact the lawyer will hardly give it a thought, and will occupy his whole time with trying to make a precise definition of what constitutes a "contract." In the same way we might expect a book on light to discuss why the grass looks green, and gold yellow, but such matters are hardly mentioned. Instead a chief part of the discussion is devoted to phenomena almost unknown to ordinary experience, the fact that light does not always go in straight lines—a thing which most go through life without consciously observing. This gives an unreal and sometimes pedantic look to the studies of the scientist, because he always seems to forget about the broad facts and to be concerned with petty details. There certainly have been workers who were open to this charge, but the great scientist is the one who recognises that what seems a petty detail is really of fundamental importance, or that some exceptional case can be included in a larger synthesis, which will take all the ordinary facts in its stride, so that there is hardly a need for him to consider the ordinary things. It is in this way that it has come about that more importance attaches to the colour of soap-bubbles, and peculiar rings that appear on certain photographs, than to such things as the possibility of aiming a rifle by looking through sights, or the way the electric light works. These last are just as important, but they are comparatively straightforward, and so we can pass them by, knowing that they will be easily understood when once we have mastered the curious phenomena which require a deeper insight for their explanation.

It will be apparent from the train of thought developed below, that the subject gives much occasion for philosophic speculations. As far as may be, I have avoided such speculations, partly because I have no skill in them, but also because a warning against them is given by the history of modern physics. It was not the philosopher Kant, but the mathematician Gauss who first discovered that the three angles of a triangle might not add up to two right angles. It was the physicist Einstein who gave life to the principle that our theories should not be concerned with things that are unobservable. As we shall see, a similar and even more fundamental idea has been introduced in the new conception of matter, and again it has been the physicists who have shown the way. There are too many alternatives offered in the undirected speculations of the philosophers, and we may expect that it will be the physicists, with the possibilities of testing alternatives by the appeal to experiment, who will continue to lead the march of our knowledge of the physical world.

CHAPTER I

THE STARTING-POINT

We are going to make a model of the world that is very different from that of ordinary life, and must start by considering what we are to use in doing it. All our conceptions are ultimately derived from common experience, so that we must start with the ordinary things of life. We can classify the things that will go to make our structure in three categories, the building materials, the building site, and the principles of the architecture. By the building materials I mean what we call matter, by the site space and time, and by the architecture the rules of mechanics to which the structure must conform. It will be best to start by a description of the general state of ideas on these subjects before the new discoveries were made. In the later chapters much of what I am going to describe will be changed, but nevertheless it can hardly be called wrong. The new discoveries have consisted not so much in destroying the old theories, as in supplementing them by an alternative point of view; this removes the inconsistencies in the old theory, and now that the reasons for its failure have been mastered, it remains an instrument admirably suited for the description of many things in nature.

We are first to consider the building materials, and shall find that ordinary matter is constructed out of two types of ultimate things called "electrons" and "protons." But there were a number of intermediate stages in this discovery, and we must trace shortly the development of the theory. Our starting-point must evidently be the ordinary matter in bulk of which we have direct cognizance. A piece of matter can be dissected again and again into ever smaller pieces, and the process may go on down to the point where we cannot get a knife fine enough to cut it, or a microscope powerful enough to see the pieces. But this does not exhaust our powers, for X-rays can be used as a sort of indirect microscope (in a way which will be described later), and they clearly show that matter is granular in structure, and the grains are the atoms. It is less than twenty years since this direct evidence was found, but we must remember that chemistry had provided indirect evidence long before.

During the last one hundred and fifty years the chemist has been patiently sorting out the different kinds of matter, showing how one kind could be constructed from other simpler ones, and has finally arrived at about ninety different *elements*—hydrogen, oxygen, copper, and so on. These are the substances which defeat him, in that he cannot reduce them to anything simpler. But he found more than this, for the other substances, the compounds, consist of two or more elements joined together in definite proportions which are always the same. The easiest way of understanding this was suggested by Dalton. He supposed that the elements consisted of atoms, and that the compounds were made by the union of two or a few more atoms of the constituent elements. A great part of the science of chemistry consists in working out the consequences, first explaining the structure of simple things like common salt, and working with marvellous skill and insight up to the extremely complicated substances which compose living matter. With this we have no concern here, as we are studying analysis not synthesis, but I have had a very definite purpose in recalling these matters, which will bring us closer to the subject of our ultimate building materials. For exactly the same argument, by which it was inferred that the elements consisted of atoms, could also be applied to electricity. This is the result of Faraday's work on electrolysis, which showed that when the atoms of a substance are electrified, as they are in ordinary solutions, their electric charge is the same in all cases, or at least is a simple multiple: one, two, three, or minus one, two or three, of a single universal charge. It is a little curious that the physicists of the nineteenth century did not make much use of the idea of atoms of electricity, even though it was suggested several times. It would not have been unnatural for the chemists to have called electricity an element like the others, but as things turned out, it is a good thing that they did not do so.

The indirect evidence of chemistry, though it made the existence of atoms practically certain, was incomplete because it did not show how small they were. It is really only in the present century that evidence has accumulated establishing their size. It would take too long to describe the various methods used in doing this —one of the most direct is the method of X-rays, to which I have referred—but all agree in giving very precise values to the weight and size of atoms. They are, of course, very small, far transcending anything of which we can get a direct conception, and my own opinion is that it is generally best to leave it at that, and not attempt to form any sharper idea of how small these things are. The use of ordinary arithmetic imposes far less strain on the mind than the attempt at a direct picture, for it is very easy to write either six or seven zeros after a number, but quite another thing to *feel* the difference between a million and ten million. The mathematician has no more feeling for this difference than the ordinary man; his only advantage is that he knows how to get on without it. But we sometimes want to consider simultaneously two things, both of which are unimaginably small, and then we do need to know whether they are the same size, or whether one is much greater or much smaller than the other, and to describe such relations a rough scale model is useful. On the whole, the most convenient scale for the general description of atoms is to multiply all lengths by ten thousand million, or, as the mathematician would say, to use, instead of the centimetre, the length 10^{-10} cm. as unit. On this scale a drop of water would have radius two or three times that of the earth. A piece of solid matter would be composed of atoms with their centres at distances of about two metres (or yards) apart. If we take a cubic centimetre of solid matter, then the number of atoms in it is something like ten thousand million million million, a number far beyond anything that can be appreciated directly, even by writing one and twenty-two zeros after it, but which is admirably expressed in the mathematical notation 10^{22}.

It may perhaps be of interest to mention the various magnitudes according to their actual sizes and without the scale change. The most fundamental of them is called Avogadro's number; it is the number of atoms in a "gramme-atom" of any substance. It is $6{\cdot}1 \times 10^{23}$. The exact distances between atoms depend on what substance we consider, but are roughly the same for all solids. For rock-salt the centres of the atoms are all placed at the corners of cubes, and the edge of the cube has length $2{\cdot}8 \times 10^{-8}$ cm. The atoms are packed together rather tightly, so that each has a radius of a little more than 10^{-8} cm. Gases are about a thousandth part as dense, and so the average distance between atoms is about ten times as much, but in a gas the atoms are moving about irregularly, instead of being arrayed in a pattern, and there is a good deal of free space between them. We may also record the weight of the atom. A hydrogen atom weighs $1{\cdot}66 \times 10^{-24}$ gr.

One consequence must be mentioned of the enormous numbers of the atoms which constitute matter-in-bulk. This is that probabilities become practically certainties. If I toss a coin ten times there is a certain small chance, about one in a thousand, that it will show heads every time. If we ask for the prospects of much more than ten successive heads, the chances against getting them become enormous very rapidly. Thus it can be shown that if a coin had been tossed forty times on each day since the time when the earth split off from the sun, it would be rather unlikely that on any single day the coin would have come down heads every time. If this is so for an experiment concerned with forty events, we can see that with the enormous numbers of atoms concerned in the experiments of physics, the improbable cases can be entirely neglected. It is likely that but for this we should never have made much progress in physical theories, for when matter is studied in bulk the effects observed are always average effects, which have been simplified to manageable proportions by the smudging out of the finer details. But to counterbalance this advantage, there is the disadvantage that the smudging made it almost impossible to work out the mechanics of single atoms from the observation of the averaged effects of a great many of them. So it came about that until the behaviour of separate atoms could be studied (and this was not before the present century), physical theories could be constructed which appeared nearly perfect, in which each atom was supposed to behave like a little piece of matter-in-bulk. In only a few of the phenomena known to the nineteenth century was there any indication that all was not well with the simple extension to atoms of the mechanics of gross matter.

In the further analysis of matter we now part company with the chemist, because his methods—heating, dissolving, and so on—are much too gentle to do more than scrape the surface of the very stable structure of the atom. With the discovery of X-rays and radium the physicist had at his disposal much more powerful means of attack. These consisted of projectiles of various kinds, which could penetrate right into the atom and knock things out of it. The first success was due to J. J. Thomson, who established firmly the existence of the *electron*, which is one of the ultimate types of material, and indeed is that one the properties of which will chiefly engage us. Thomson showed that whatever the method used in attacking the atom, and whatever the type of atom attacked, the only thing that ever comes out of the atom is one or more electrons. These are all identical with one another; they all carry the same electric charge, and are of the same weight, no matter whether they are extracted from oxygen or gold or anything else. Moreover, the charge can be identified with the electric unit of charge as found by the chemistry of electrolysis, to which we referred above; in electrolysis each atom either gains or loses one or more of the electrons which it normally contains.

The charge on the electron is negative, but this is only an accident; the choice of what was to be positive electricity and what negative was originally quite arbitrary, and it happens to have made the electron negative. Its mass, or weight, is very much less than that of any atom, though not quite insignificant in comparison. If we take 1 for the mass of the electron, then the lightest atom, hydrogen, has mass about 1845, and the heaviest, uranium, is 238 times as heavy again, or about 440,000 times as heavy as the electron. Until the advent of the new theories, the electron was regarded as a little particle of electricity; on grounds which need not be described, its size was estimated and found to be about 10^{-13} cm. On our scale model, with an atom about two yards across, this would represent a length of a hundredth of a millimetre, so that an electron would be hardly visible to the naked eye, but could easily be seen in a microscope without using a high magnification. This rather materialistic way of picturing the electron will be very much modified later, as we shall see. The occurrence of electrons in every type of matter whatever shows that they are one of the basic constituents of nature. We will now leave them and explore the atom further to see what other constituents it may contain.

The discovery of the other main constituent, the *proton*, took place in two stages, the first being the discovery of the *nucleus*. Rutherford deduced its existence from the behaviour of the rays emitted by the radioactive elements. These rays are of three kinds, called α, β, and γ, and all are interesting. The β-particles [1] are simply electrons going at enormous speed, while the γ-rays are like light and X-rays. Here we are concerned with the α-particles. These have mass four times the mass of a hydrogen atom, and are really atoms of helium, the next lightest element after hydrogen. They only differ from helium in that they carry an electric charge, which is positive and twice as large as the negative charge of the electron.

---

[1] The terms β-particle and β-ray are used almost indiscriminately, though strictly speaking one would say that the β-ray is the path of the β-particle. The same applies for the α-rays. The term γ-particle is not used.

In fact the ultimate fate of an α-particle, after it has done its work, is to pick up two electrons, become neutral, and turn into an ordinary atom of helium gas. The α-particles are projected out of radium atoms at enormous speed, some at a fifteenth of the speed of light, and it is in the period before they have lost this speed that they produce the effects which concern us.

The α-particles go straight through everything in their way, gradually losing speed as they go. Through air they go two or three inches, and it is easy to exhibit their tracks, because a cloud can be condensed on the atoms which have been churned up by an α-particle Solids are, of course, much less pervious to them; an α-particle can go through a thin sheet of paper, but not through a thick sheet. Now we know that the atoms of solids are packed very close together, so that there is no room between them, and it follows that the α-particles must go right through the atoms themselves. Here is a powerful probe to show what the atoms contain. The first thing found is a considerable number of electrons, but these are so much lighter than the α-particles that they hardly affect it. But every now and then an α-particle is given a large deflection by an atom, and this can only be explained dynamically by supposing the atom to contain a nucleus of small size, which carries practically the whole mass of the atom. In Plate I, A an example can be seen of an α-ray which has been strongly deflected by the nucleus of an oxygen atom.

The detailed properties of the nucleus were worked out partly by means of the α-particles and partly by other indirect means. It was shown that the elements could be arranged in numerical order, so that hydrogen is No. 1, helium No. 2, . . . copper No. 29, right on to uranium No. 92. The atomic number, as this is called, describes the electric charge on the nucleus, measuring it in multiples of the charge on the hydrogen nucleus, which is equal to that of an electron but positive instead of negative. An ordinary atom consists of this nucleus, neutralised electrically by the appropriate number of electrons. In addition to its electric charge, the nucleus carries practically the whole mass of the atom; for the electrons are by comparison so light that they barely make a perceptible addition to the weight.

Before taking up the analysis of the nucleus, it will be well to digress and consider the general structure of the atom. Here again it must be said that the picture we draw will be modified later when we have gained the new point of view. It is the picture that would have been accepted five years ago, and it still remains very useful for many purposes. Consider our scale model in which a drop of water became as large as the earth, and an atom became a couple of yards across. Let us say we are considering element No. 29, which is ordinarily called copper by those who have no taste for red-tape. At the centre of the atom there will be the nucleus, perhaps just visible to the naked eye. Let us suppose that this nucleus is made of the same material as the actual nucleus, that is to say, that we take real nuclei and jam them together until they fill up the volume given by the change of scale. I am not very sure whether this process really helps our ideas, but if we do it we find that the barely visible nucleus would weigh about a hundred tons. This nucleus is surrounded by 29 electrons, each of which weighs about 2 lbs. They are scattered over a region about a yard in radius, but are mostly only a few inches from the nucleus. It is not desirable to be much more precise than this, for, as I said, the whole picture is very substantially modified by the new theory.

We must now turn to the question of the structure of the nucleus, the evidence as to which is mainly rather indirect. It comes from measuring the weights of atoms. The methods of chemistry depend essentially on the electrical properties of the surface of the atom, and are not delicate enough to bring out the effect of a small change of weight; the chemist is studying the properties of atomic number only, and his atomic weight is the average value of the weights of all atoms possessing the same atomic number. By skilfully devised electric and magnetic experiments Aston succeeded in measuring the weights of individual atoms, and he found that most of the elements, which appear perfectly uniform to the chemist, are really mixtures. He showed that copper, for instance, is really composed of two "isotopes," one weighing 63 and the other 65 times as much as a hydrogen atom. He has investigated the majority of the elements, and has found that in all cases the weight of the atoms is nearly an exact multiple of the weight of a hydrogen atom.[1] It is natural to draw the conclusion that the nucleus is built up out of hydrogen nuclei and electrons. The atomic weight shows how many hydrogen nuclei there are, and the charge depends on a partial neutralisation of the charges of these nuclei by electrons. It is a matter of simple arithmetic to calculate the numbers of either in any nucleus. Thus a copper atom may have weight 63 or 65, and so may contain either 63 or 65 hydrogen nuclei. Its atomic number is 29, so that in the former case it will have $63-29=34$ electrons, and in the latter $65-29=36$. Calculations of this kind can be made for every isotope of every element. No very precise rules for the numbers have yet been found to describe the relationship between atomic number and atomic weight, but broadly speaking the atomic weight is rather more than twice the atomic number, so that the number of electrons in the nucleus is apt to be rather greater than the atomic number. In view of the universality of the occurrence of the hydrogen nucleus, it is desirable to have a name for it which should not so definitely connote its connection with a particular chemical substance, and so it has been renamed the "proton," the *primitive substance.*

---

[1] Closer investigation shows that the multiple is hardly ever exact, but the reason for the difference, though very interesting from another point of view, is more or less understood and need not delay us.

The existence of protons is obvious merely from the existence of ordinary hydrogen, but so far the evidence for their universality is only indirect. It is very like the old chemical evidence for the existence of atoms, for both depend on arguments from the constancy of proportions. When the chemist has discovered the composition of a substance, he sets to work to synthesise it out of the component elements, and we should like to do the same, for example, by taking 63 protons and 34 electrons and turning them into a copper nucleus; but this cannot be done. Since the direct proof is lacking, the next best thing is to know why, and this knowledge we can claim; we cannot expect to perform the synthesis until we can produce temperatures of a great many million degrees. That seems an adequate answer to the challenge that we have not done it, and if a sceptic points out how weak the evidence is, the right answer is to tell him to go to a place where the temperature is a great many million degrees.

It is nearly always easier to destroy than to create, and though we cannot create nuclei out of protons and electrons, yet we can sometimes break up existing nuclei or can witness their spontaneous explosion. The case most studied is that of the radio-active substances. As we saw, these sometimes give β-particles, which are electrons, and this type of disintegration shows that there are electrons in the nucleus. The α-particles, the other chief type of emission, are, it is true, not protons but helium nuclei, and from this we infer that the nucleus of radium is composed, at any rate in part, of protons arranged in quartets as helium. There is good reason to believe that four protons and two electrons combine together into a specially stable "compound," and that the nuclei of most of the elements are largely constructed by combining together much more loosely these quartets along with a few extra protons and electrons. Additional direct evidence that the nucleus contains protons is furnished by the artificial disintegration which Rutherford has succeeded in producing. When a very intense beam of fast α-particles is sent through some of the lighter elements, a few of them can penetrate right into the nuclei of those elements and break them up, with the emission of a proton. An example is shown in Plate I, B. This seems to complete the proof that protons really are a primary constituent of matter; we can hardly expect anything better until someone is skilful enough, and also brave enough, to make a furnace which gives millions of degrees of temperature.

We are thus led to believe that there are only two primary constituents of matter, electrons and protons. Both of them have electric charges of the same value though opposite signs, but their masses are quite different, the mass of the proton being about 1845 times as heavy as that of the electron. When we encounter a pure number among our physical constants as we do here, we anticipate that it has some very fundamental meaning, since it in no way depends on the units in which things are being measured. This will surely be the case with 1845, but no one yet has any idea of how it comes about, and it is such a curious number[1] that it does not fit into most of the simple attempts to make an explanation.

I shall add to these two primary constituents a third which is on a rather different footing. This is light, which is regarded as including wireless waves and X-rays as well as visible light. The whole history of science in the nineteenth century regarded light as something radically different, a wave motion of a medium instead of a grain moving through a medium. But during the eighteenth century it was regarded for the most part as corpuscular, and the twentieth tended to return to a similar opinion. At the present time a union of the two opinions has come about, so that both light and electrons and protons are held to have properties both of the type that the nineteenth century recognised as those of light and those of matter. This, however, is the principal subject with which we shall be concerned throughout this book, so we pass on. Light corpuscles are sometimes called "photons," atoms of light. It is an unfortunate accident of the Greek language that two of our three elements, protons and photons, each very happily named when considered by itself, should sound so alike. For this reason it is often better to refer to them by an older name, "light-quanta"; we shall use the words photon and light-quantum indifferently. Photons differ from electrons and protons in that they can never stay at rest but are always travelling at thirty thousand million centimetres a second. A photon goes seven times round the earth in a second. This is often observed with the photons of wireless, for an echo can sometimes be heard a little later than the original sound; the sending station has poured photons out into space and some of these have gone round the earth before they are caught by the receiving station.[1] Another important difference is that photons do not affect one another in any way, as do electrons; two beams of light can be crossed without influencing one another, and this means that the photons can cross without colliding. But photons do affect electrons that they meet; and they must also affect protons, though the much greater weight of these makes the observation more difficult. When a photon strikes an atom it often knocks an electron out of it, or shifts it inside the atom, and this is the only way in which photons are ever observed, though the details of the observation can vary a great deal. Thus the photons may knock electrons out of a metal plate, and we may measure the loss of charge with an electrometer; or they may knock them out of the atoms of a photographic film, which we then develop; or more directly they may knock electrons out of the atoms in our own retina, so that the atoms, feeling themselves chemically imperfect, send a protest in that sense along the optic-nerve to the brain.

---

[1] Its value is uncertain to three or four units.

The habits of scientific thought of the last century have become so ingrained that it may seem surprising that we should attribute the same sort of primitive character to light as to electrons and protons. If we analyse why we feel this surprise, we find that it is because we are apt to think of electrons and protons as little grains, as I did in the scale model I used. As grains they would have a character of individuality, whereas no one wants to attribute this individuality to rays of light.

---

[1] The echo is usually about a quarter, instead of a seventh, of a second later than the first sound, because the waves are not *in vacuo*.

But the modern view, which we shall be explaining at length, justifies the likeness not by giving light individuality, but by denying the individuality of electrons and protons. It may also be said that it does not seem impossible that we may ultimately be able to do without photons altogether, explaining their apparent effects by direct actions between electrons and protons. I shall not here go into more detail about photons, as it will form the subject of much discussion later.

The second category that we have to discuss is what I called the building site. Broadly speaking, by this I mean space and time, but we can convey better the more definite connotation by using the word *medium*. In the early history of the theory of light there was a very celebrated controversy, which lasted for more than a century, between the schools who believed that light was corpuscular and those who believed it was a wave. It was a question of whether light was like a flight of arrows, or was an effect propagated through a medium permanently there. After a long lease of life the corpuscular theory was routed, as it seemed finally, by the great optical discoveries of the beginning of the nineteenth century, an epoch in optics almost as great as the beginning of the twentieth has been in atomic physics. The great discoveries of Young and Fresnel seemed to establish very definitely that light was a disturbance propagated through something like waves travelling over water, and a name, the æther, was invented for the medium carrying these waves. The question naturally arose as to what the æther really was, and much of the speculation of the nineteenth century was directed to finding this out. A great many models were tried and found wanting. At first it was thought to be like an elastic jelly, but that led to consequences which were not borne out by experiment. Then other models were tried of ever-increasing elaboration. These models may be said to have solved the problem, but not at all in the way intended. They were so complicated—things like tops mounted on universal joints were common in them—that they showed that they were useless. The æther should be one of the fundamental things in the world, and it is not to be expected that it can be expressed in terms of the things which it is itself going to explain later.

In recent times the idea of the æther has taken on a much vaguer and more satisfactory shape, merely as space endowed with certain properties. It has been defined as the "subject of the verb to undulate," and it acts as this subject very well. Its properties are not to be expressed in such crude ideas as elastic jellies, but are merely those things that we require it to do. With the advent of Einstein's relativity, it has often been said that the æther has been abolished. As far as concerns elastic jellies, that is certainly so—if they still had adherents before relativity—but it is nevertheless useful to have a subject of the verb to undulate, and relativity is from this point of view merely a new and much more fundamental quality of the æther. That is not intended to belittle relativity, for its general point of view of questioning the reality of anything unobservable is one of the greatest revolutions in scientific thought that has ever occurred. All that relativity has done to the æther is to show that, so to speak, we cannot make marks in our jelly and expect to find them again. But with the proviso that æther merely is a convenient noun to describe the properties of space, it is often a convenient grammatical construction to have.

We can impose on the æther, or space, all the new properties of which we are going to find it possessed, without straining its capacity. It is the business of a company of universal carriers to carry everything placed in their charge. They are incorrectly styled if they show themselves incompetent to carry anything in the universe; we are merely going to demand of the æther that it shall be a true universal carrier.

It remains to discuss the principles of the architecture. Here again we must link the question with things that we can understand intuitively—that is to say, with the behaviour of matter in bulk; that is to say, with ordinary dynamics. The full discussion of dynamics is severely technical, but we can fortunately get on very well without studying the technique. Our general aim in studying nature is to say, "I know that I have a set of bodies (whether atoms, billiard balls, or planets) placed in such and such places, and moving in such and such ways now; where will they be and how moving at any later time?" The greatest advance in the attack on this problem was due to Newton, who recognised that it could be separated into two parts, the first quite general, which he solved completely, and the second special to each case. He gave general rules to say where the system will be at the next instant of time; these are called the equations of motion. It seems obvious to us now that a knowledge of the whole motion can only be derived from the instantaneous motion, yet in fact this had not been the way the matter was regarded previously. Thus Kepler described how a planet moves in an ellipse, but it was Newton who cut the ellipse up into little pieces, studied the motion in each little piece, and so was led to his law of gravitation.

Though this first part of the problem of motion is really the chief one, and incidentally is what causes trouble to beginners in the subject, yet it was so completely solved by Newton that, as far as concerns gross matter, there is little more to be said about it. Interest has, therefore, been concentrated on the second half of the problem, the step from the instantaneous rates of change of the motion to the description of the motion at a finite time later. This is a matter that is special to every problem, but there exist certain principles which are extremely helpful in the solution, principles so general that they may in a sense be regarded as fundamental laws. These are called the conservation laws.

The first of these frequently conserved quantities is *momentum*. The momentum of a body is technically defined as the product of its mass and its velocity. Newton's Third Law asserts that "to every action there is an equal and opposite reaction," and in spite of its different wording, this implies the conservation of momentum—that is to say, that when two bodies A and B interact, what A loses in momentum is gained by B. Anyone who has fired a rifle without holding it quite tight enough will know that the momentum of the bullet causes an opposite momentum of the rifle that can be quite painful. The first of our laws then is that in any isolated system the total momentum is constant. Thus if there is a collision, say between billiard balls, and I see how one of them moves after the collision, I can confidently deduce the motion of the other from the fact that their joint momentum will be the same as their joint momentum before the collision; it is true that they are not completely isolated from other influences, such as air resistance and friction of the table, but during the short time of the collision the effects of these will be insignificant. The conservation of momentum is a specially useful rule in connection with collisions, using the term in the widest sense. For example, light has momentum, though not much of it, and so every time a photon is reflected at a mirror, this momentum has to be reversed, and the mirror will be pressed backwards in doing it.

The second important conservation law arises in a different case. If I whirl a stone round at the end of a string, keeping some slack in the hand, and then pay out the slack, the stone will go slower as the string gets longer. This illustrates the conservation of another quantity, angular momentum, which applies especially for motions round a centre or axle. The angular momentum of the stone is technically defined as the joint product of its mass, speed, and of the length of the string. As the angular momentum has got to stay constant, when the string is lengthened the speed must diminish. The most important example of angular momentum is that of the planets round the sun. A planet is retained in its orbit by the direct pull of the sun, which cannot affect the cross-wise motion, and so the planet retains its angular momentum constant. As a fact, of course, the planets attract one another a little too, so that they do affect one another's angular momentum, but what one gains the other loses, and the sum is constant. The total angular momentum of the solar system is thus absolutely invariable no matter what interchanges there may be between the planets. The earliest attempt at a theory of cosmogony was Kant's nebular hypothesis. Kant was a very great philosopher, but he was weak in dynamics, and his theory failed from the very start, because he supposed that collisions between the particles of the nebula could generate rotation in the solar system, and this is contrary to the principle of angular momentum.

There is one more conservation theorem, and it is much the most important of all. This is the conservation of energy, and it is derived from the engineering experience that it is impossible to make a machine work without providing it with power. Translated into more appropriate language, we say that the energy of an isolated system can never increase. The most primitive kind of energy is *kinetic*, derived from motion. The kinetic energy is proportional to the square of the speed (whereas momentum was proportional to the speed itself). When I throw a ball up its energy remains constant, and as it slows down the kinetic energy is turned into *potential* energy, and then, as it falls, back into kinetic again. In many ordinary motions the energy of systems diminishes on account of friction, and one of the great discoveries of the nineteenth century was that friction generates heat in exact proportion to the amount of energy disappearing. The obvious deduction is that heat is a third kind of energy, and that we can maintain the conservation by adding this on to the kinetic and potential. Heat is nothing but the random motions of all the atoms, and so all that is happening when energy seems to disappear—as when a rolling ball comes to rest—is that what was a systematic motion in bulk has turned into a disorganised motion of the atoms.

As we have seen, our only intuitive method of understanding atoms is to extend to them ideas which we can appreciate from matter in bulk, and so we naturally extend to the atoms the principle of conservation of energy. But we have to make one most important difference, for we are going to think about the motions of individual atoms as their kinetic energy, and so there is now no room for the idea of heat. For matter in bulk heat was regarded as the kinetic energy of the atoms; and so, when we come to consider the energy of individual atoms, heat has no meaning, and we are only allowed to have kinetic and potential energy. In some ways this makes the consideration of atoms easier, since one of the possible forms of energy is excluded; but this exclusion imposes difficulties too, because we can no longer escape the rigorous enforcement of the rule that the energy to be conserved may only be of the two kinds, kinetic and potential.

There is no possibility of dealing with modern physical ideas without some understanding of energy and momentum. We have described some of their properties shortly; and it is rather important to remember the distinction. Both depend on the speed of motion, but energy more than momentum, since it is proportional to the square of the speed. In the early days the technical terms were much less fixed than they are now, and a name was used for kinetic energy which gave a very good description of its character. The Latin word *vis* was used for a force, that is the thing which tends to move a body. But when *vis* was qualified by an adjective it meant something which we should not now admit to be of the nature of a force at all. Energy was called *vis viva*, and the obvious contrast would be *vis mortua* for momentum. Though I can find no authority for this last term, the distinction between "live" and "dead" is a very good loose description of the difference between energy and momentum.

CHAPTER II

WAVES

In the present chapter we shall study the general behaviour of waves. We shall be dealing with matters which have been fully understood for the last sixty years at least, and yet of which the bearing on our present subject was only perceived about six years ago. Why is it that among the innumerable phenomena of nature the physicist has selected the waves of the sea as being such an important type? The reason is that he is largely concerned with the methods by which causes in one place produce effects in another, and the propagation of water-waves is one very familiar type of method. To most of us it is intellectually unsatisfactory to accept "action at a distance," because this only amounts to saying without further inquiry that a cause in one place can produce an effect in another, whereas we know that in a great many cases we can see the effect travelling. The most elementary way in which I can attract anyone's attention is to throw a stone at him, or as the text-book might put it, to project a particle so that its trajectory passes from A (me) to B (him). This we shall consider later. Another way is to poke him with a stick and this has quite a different character, because there is no transfer of matter from me to him—a small motion that I produce in the wood at my end turns into a small motion at his end. Here there is a transfer of motion without a transfer of matter, but wood is so stiff that it would require rather accurate time-measurements to show that the motion at my end takes place a little time before that at his. In making a picture of the process we naturally take an example where we can see everything happening in detail, and there is none so familiar as the propagation of waves over the surface of water.

All processes of this kind in nature are called waves, and we shall have to consider some of their general properties, but it will be best to start with certain specially simple types. Waves on the surface of water usually have the characteristic that their crests are spaced at fairly regular intervals, and that the waves travel in a direction at right angles to the lines of the crests. Most actual waves in the sea conform only roughly to this description, but it is good enough to justify us in beginning with the study of this simplified case. Consider the motion of a perfectly regular progression of waves, travelling over deep water at right angles to the line of a crest. The profile of a vertical section of the water taken across the crests would be the sinuous curve called a sine-curve. Such a wave is called a harmonic wave, and its characteristic is that it has crests and troughs regularly spaced and extending indefinitely in both directions. The details of the motion can be solved mathematically, and the results are as follows. Each particle of water moves round and round a circle with fixed centre. Take a particle at the surface. When the crest is passing this point it is at the top of its circle and moving forwards. After the crest has passed, it starts moving down and is going vertically downwards at the moment the water is at its mean level. At this instant that particle is as far in front of its average position as it will ever be. At the depth of the trough it is moving backwards horizontally. Circles have been drawn in other places in Fig. 1 to show the paths of the particles there; the arrow marks where the particle is and shows the direction in which it is moving. The particles of water below the surface also describe circles, but of smaller radius, and as we go down the radii diminish rapidly in size. At a depth equal to the length of the wave from crest to crest the motion is only a five-hundredth of that at the surface.

The study of water-waves can serve to introduce certain technical terms which we can hardly avoid using. What characteristics are needed in order to describe the wave completely? First there is the *wave-length*, the distance from crest to crest—this needs no explanation. Then there is the speed with which the wave is travelling. We take a crest and see how many feet it goes in a second; this is called the *wave-velocity*. Associated with these two is a third which is quite as useful. If we know the length between crests, and how fast the crests are travelling, we can evidently work out the interval of time between the passage of successive crests past any point, or alternatively the number of crests which pass the point in a given time. This last is the more useful quantity. The number of crests passing a point in a second is called the *frequency*. Another characteristic of the wave is its height. This is always taken to mean *half* the height from crest to hollow, that is to say, the greatest elevation or depression of the surface above or below its

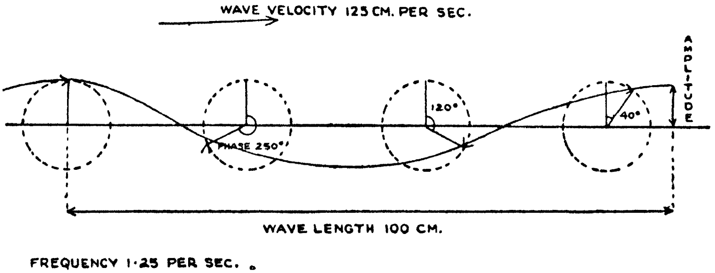


FIG. 1.—HARMONIC WAVE

The terms necessary for its description are marked in the diagram. The speed given is correct for waves 100 cm. long on deep water.

average level. For water-waves the name height is correct and suitable, but we shall be dealing with other kinds of waves too, and a more general term is therefore convenient, so that we shall not be committed to thinking about any special kind of waves. We therefore call the height (or rather the half height) the *amplitude.* The technical terms used in describing a harmonic wave are shown in Fig. 1.

If we know wave-length, frequency, and amplitude we have a complete description of the characters of the wave, but we do not know exactly where the crest is at the instant considered and for this we must introduce one more quantity, called the *phase.* If we take some particular point and say that at some particular instant the crest is passing that point, then we shall have a complete specification of the wave, and can calculate from the wave-velocity, etc., where it will be at any future time. To give the phase at a point is to specify what part of the wave is passing that point; for example, the water might be rising but still below its mean level. The actual numerical description of the phase is best done by consideration of the circles the particles describe. We measure the angle from the top point, and call this angle the phase at that place. Thus the crest is at zero phase, the trough at 180°, and half way between we say the phase is 90° if the wave is falling and 270° if it is rising. At any point the phase goes steadily round the 360°, but if we consider a different point, the phase, though also going round, will at any instant be different from that of the first point. Thus at the place B a quarter wave-length *in front* of A the phase is 90° *behind*, or in other words is 270°, since B has to wait for the passage of a quarter of a wave before it gets to the phase which A originally had. It is always hard to remember the exact meanings of technical terms, and I fear these ones may have been tiresome. Fortunately we can usually take them quite loosely, so that when we talk of phase we only need remember that a wave is a cycle of events, and that we are concerned with the fact that at any instant of time different places are in different parts of the cycle.

I have described rather fully the character of water-waves, because it is so familiar, but the same terms are used for other kinds of motion of the harmonic type. Thus a stretched wire like a telephone wire can carry waves along it, as we may verify by tapping one end and seeing the waves travel. In this case a particle of the wire moves sideways to and fro at right angles to the length of the wire. If the wave is a harmonic wave, the same terms, wave-length, etc., apply to it; the amplitude now means the maximum distance which a particle moves sideways from its mean position, and we can speak of phase just as before, even though the particles of the wire do not actually describe circles. Or take our original example of poking with a stick. In this case the stick gets compressed by the thrust at one end, and this compression communicates itself to the next piece, while the first part relaxes and a wave of compression travels down the stick. Here we can no longer so easily make a drawing to show what is happening, since the stick remains quite straight, but nevertheless all the same ideas apply. If we thrust at regular intervals of time a harmonic wave of alternate compression and extension will travel along. The distance between two successive compressed places is the wave-length, the rate at which they travel is the wave-velocity. The amplitude can be described either by the greatest thrust that is propagated, or alternatively by the very small displacements of the particles of the stick which that thrust produces. The phase again describes what part of the cycle of changes is occurring at any point of the stick at the instant under consideration.

Other and more important examples are furnished by sound and by light. In sound the air is compressed and then released, and, as in the stick above, waves of alternate compression and rarefaction travel out through the air. Here again we use the same terms wave-length, etc., and among them we may recall that frequency has a very intimate meaning for us since the frequency is the pitch of the musical note. Indeed the name "harmonic wave" is borrowed from the theory of sound. The amplitude determines the loudness, and can be measured by the greatest amount of compression of the air in the wave. It may seem strange that I should speak of light as waves here, whereas before I grouped the photons of light along with electrons as particles. It is typical of the new outlook that we have to face this duality, and we shall be devoting much attention to its explanation. For the present I am speaking from the point of view of the nineteenth century, when light was only a wave and nothing else. From this point of view the frequency of light determines its colour, and the amplitude its brilliance, but we shall have so much to say about this later that we may pass on for the present.

When we are studying the motions of water-waves we observe them comfortably (or perhaps uncomfortably) from the deck of a ship, and thus, looking at them sideways, we can see how high they are—that is to say, we can directly measure the amplitude. The same is true for the vibration of a wire, but for sound it is not so. When we observe an air vibration we do not see a wave going sideways past us and look at the motions of the air particles; we have the wave coming right to our ears and producing an effect in their mechanism which we call noise. The mechanism is called the fibres of Corti, a little keyboard of hairs each of which vibrates to some particular note.[1] The effective amount of its response is not proportional to the amplitude, as can be demonstrated from mechanical principles, but to its square. Thus if we halve our distance from a source of sound, the amplitude is doubled, but the sound is four times as loud. The loudness of the sound is technically called its *intensity*, and is a much more important thing than the mere to-and-fro motion of the air. The intensity of a sound is a measure of the *energy* that is being propagated through the air. In the same way the intensity of water-waves is proportional to the square of their height, and most of the effects of the waves depend on their intensity. Thus when we watch the waves from a ship, we see their amplitude, but our other sensations will be more closely related to their intensity; waves that the bad sailor sees to be twice as high will make him four times as unhappy.

We have so far merely described the various qualities of a single harmonic wave, without considering how they are related together. Much the most important principle is that the amplitude has no effect on the wave-length or wave-velocity. A loud sound, which has large amplitude, goes no faster than a faint sound; and great water-waves travel at the same speed as little ones, provided that the wave-length is the same. To avoid misapprehension I ought to say that this ceases to be strictly true for the great waves of a storm at sea; its validity depends on the height being a small fraction of the wave-length, which is certainly not true in a rough sea. But it is true in a large number of important cases that the wave-velocity in no way depends on the amplitude—for example, for water-waves of small height, for sound, except the most terrific explosions, and as far as we know universally for light. This general principle is a great simplification, because we need never think about the amplitude when we want to know the wave-velocity. There are no such simple relations between wave-velocity and wave-length; every case must be considered by itself. Thus for light the wave-velocity is the same whatever the wave-length, and this is also true for sound. On the surface of deep water long waves go faster than short; for example, waves a metre long go at a rate of a metre and a quarter a second, but waves four metres long go at two and a half metres a second. On the other hand, in some cases short waves may go faster than long; this is so for the transverse vibrations of a bar. Every medium must be the subject of a special investigation to determine how the wave-velocity is related to the wave-length.

---

[1] This is perhaps an old-fashioned theory of audition. Modern work suggests that the matter may not be so simple, but that cannot affect the mechanics of the process.

The independence of the amplitude from the other characters of the wave is a special case of a very important principle, *the principle of superposition.* This asserts that a medium may carry two sets of waves at the same time, and each will be propagated just as though the other were absent. At every place the resultant amplitude is got by simply adding together the amplitudes of the two waves. Thus if we take two identical waves and superpose them we shall simply get a wave just twice as high, and this will be propagated at the same rate as the two parts of it. Or we may superpose the waves with a difference of phase as in Fig. 2; the resultant will be a third wave of a new amplitude and new phase, but going at the same speed. A particularly interesting case arises when the phases differ by 180°, for that means that the maximum of one falls on the minimum of the other, and they cancel out completely.

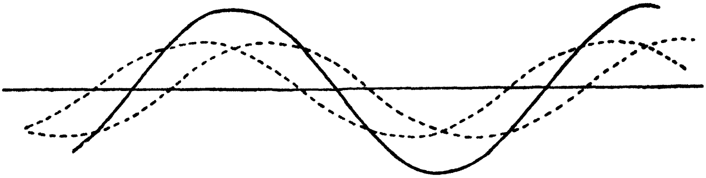


FIG. 2.—SUPERPOSITION OF WAVES

The broken lines represent two equal harmonic waves. The firm line, constructed by adding their heights together, represents the actual surface of the water.

Another example is illustrated in Fig. 3. Here two equal waves are going in opposite directions, as happens when water-waves are reflected by a breakwater. The result is that at certain places the crests meet and the water alternately jumps up high and sinks down deep. There are other places where the oncoming wave is at the top when the reflected is at the bottom and the water stays permanently at its mean level. Such a place is called a *node*. The general motion is called a standing wave because there is no progressive motion; the crests rise and fall in the same places midway between the nodes,

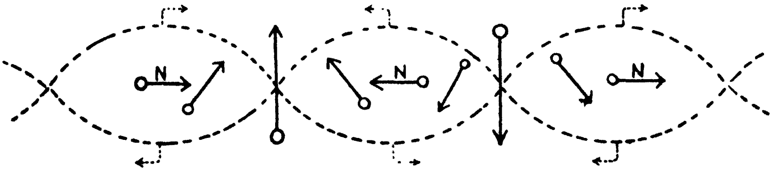


FIG. 3.—STANDING WAVES

Two equal waves are travelling in opposite directions, their motions being indicated by the broken arrows. At the moment shown the surface would be flat. The firm arrows indicate how the particles of water move. At a time a quarter of a period later the surface will correspond to the heads of the arrows, at half a period it will again be flat, and at three-quarters of a period it will correspond to the tails of the arrows. The *nodes* are marked N.

and the water particles no longer describe circles, but straight lines.

The *principle of superposition* is of enormous importance to our subject, and it is interesting that it was historically the very first argument ever used to show that light must be a wave-motion. Huygens maintained this view on the ground that if light were not like a wave, but like a flight of arrows, then sometimes there would be a collision when two beams of light cross one another. In that case one would expect that light would be scattered from the place of intersection of the beams, which in fact does not happen. On the other hand, according to the principle of superposition, each beam will travel just as though the other were not there.

Hitherto we have dealt only with the kind of waves called waves of one dimension—that is to say, our water-waves were supposed to have infinitely long straight crests at right angles to the direction of their motion. We are now to study the way in which waves behave when they are able to spread out sideways. Suppose that we have a narrow canal opening into the sea as in Fig. 4, and that regular harmonic waves are travelling down the canal. What will happen when they reach the sea? Obviously they spread out into circles round the mouth and travel away as circular waves, something like the ripples made when we throw a stone into the water. The distance between the circles representing the crests is the same as the distance between crests in the canal, and they travel away at the same pace, but the amplitude decreases as they go; we need not consider exactly how. Next suppose that instead of a canal we merely have a long wall across the sea, with a narrow gap in it, and that harmonic waves are travelling towards this wall on one side. They will be reflected by it everywhere except at the gap. On the first side except near the gap there will be set up standing waves of the kind we described, but these do not concern us at the moment. On the second side we shall get just the same circular wave travelling onwards.

Let us now suppose that we have two gaps in the wall. This gives us the mental picture we want for the immensely important phenomenon called *interference*. Each will give a circular wave, and the principle of superposition tells us that the resultant motion of the water can be found by simply adding up the two separate motions. To see what happens we make the diagram of Fig. 5, but instead of merely marking the crests, we draw the profile of the water for various directions. Near the two openings *A* and *B* there is a rather complicated and confused state of affairs, but we need not consider it. At a great distance, however, at *P*, which is equidistant from *A* and *B*, the waves will clearly reinforce one another, but at another place *Q* we see that the crest

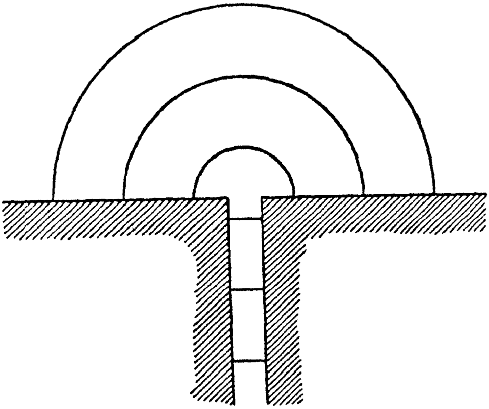


FIG. 4.—WAVE EMERGING INTO SEA

A harmonic wave travels along a canal opening into the sea. The wave spreads out into semicircles. The distance between crests in the sea is the same as in the canal.

of one will fall on the trough of the other and the two waves will thus destroy one another. For this to happen *Q* must be half a wave-length nearer to *B* than it is to *A*. At *R* the crests again fall together, because the wave from *A* has taken exactly one wave-length more to get there. The total result will be that there will be quiet strips of water and between them stormy strips. An example is shown in Plate II. The phenomenon is called Young's Experiment; it was this great physicist who first succeeded in exhibiting it for light.

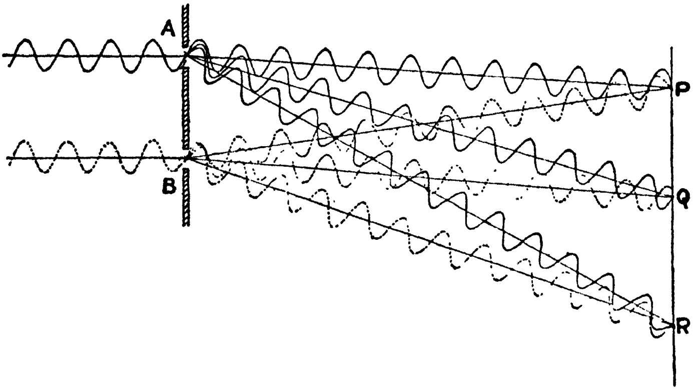
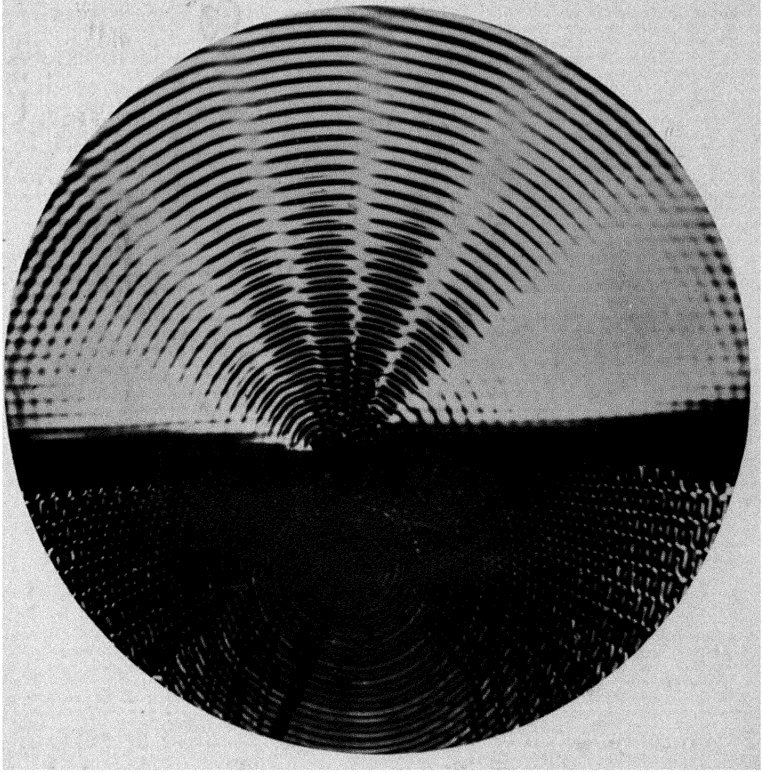


Fig. 5.—Young's Experiment

A wave passes through the two gaps *A* and *B*. *P* is equidistant from *A* and *B*, and the waves reinforce one another there. *Q* is half a wave-length nearer to *B* than *A*, and the waves, being there in opposite phases, cancel out. At *R*, which is a whole wave-length nearer to *B* than *A*, the waves reinforce one another again.

This phenomenon, as we have said, is called *interference*, and is of quite first-class importance. We have the surprising result that in some places the effect of two gaps is not, as we might anticipate, twice as great as that of one, but is nothing at all. In between, where the waves add together, the amplitude is twice that for a single gap, and therefore the intensity, which is the important quantity, is four times as great. Averaging over all parts of the sea we get for half of it four times the intensity

Plate II



Young's Experiment

Illustrated by an instantaneous photograph of the surface of a pool of mercury The pool is crossed by a barrier in which there are two small gaps close together. Waves are set up by a tuning-fork touching the surface. The centre of the circular waves generated by it can be seen in the lower half of the figure. In the upper half the waves emerging from the gaps *interfere* and cause the surface to be quiet and disturbed in alternate bands.

due to a single gap, and for the other half zero; so altogether we get on the average double the effect from two gaps that we get from one, which is what we should expect.

I have described the phenomenon of interference for water-waves, but it applies equally well to any kind of wave in two dimensions or in three. The two most important types are sound and light, and in neither is it easy to observe. The difficulty in sound is that the wave-lengths of ordinary sounds are rather large, several feet, and consequently the experiment has to be done on an architectural rather than a laboratory scale. Moreover, even so, it is not easy to observe the places of silence, because sound is so easily reflected by walls and ground, that even if the direct wave is absent there are usually echoes coming from somewhere to the observer. Again, the experiment must be done with a source giving a pure note, for the places of interference depend on the wave-length, and so are different for sounds of different pitch. If one looks for interference with a noise composed of a number of different tones, though some may be absent, the others are not, and the effect is hardly noticed. All these difficulties explain why the interference of sound is not a familiar thing, but of course they can be overcome, and by using a pure tone of high pitch (and so of short wave-length) in the open and away from buildings the interference can be observed.

For light the difficulty is the opposite—the wave-length is too short. The wave-length of ordinary light is about a twenty-thousandth of a centimetre. This does not convey much to us, but it may be helpful to express it in terms of the scale model we described before. On that scale a drop of water became the size of the earth, and an atom about a yard in radius; on this scale visible light would have wave-length about 5 km. or 3 miles, somewhat longer than long-wave wireless. If light is sent through a very small hole in a shutter—and it is not easy to make a hole small enough—it does not go in a straight ray in the way we should expect, but spreads out uniformly in all directions on the farther side, like the sea-wave going through the gap in the wall. A white screen will be uniformly illuminated all over, and will not show a bright spot opposite the hole. Now suppose that two small holes are made rather close together in the shutter, and are illuminated with light of a single colour, and therefore of a definite wave-length. It will then be found that the screen is no longer uniformly illuminated, but is crossed by bands of darkness; these bands show the places where the waves from the two holes arrive in opposite phases, so that they cancel. Midway between them the screen will be four times as brightly illuminated as for a single hole, since the brightness depends on the intensity, and quadrupled intensity corresponds to doubled amplitude. The discovery of the effect was due to Thomas Young in 1801; it was the starting-point of modern optics, and one of the most important experiments ever made. To give an idea of how difficult the experiment is, when the holes are $\frac{1}{20}$ in. apart and the screen is a yard away, then the breadth of each band is about $\frac{1}{30}$ in. When we reflect that this is a favourable case for the observation, it is hardly surprising that interference is not very familiar to us in ordinary life.

We will now return to the consideration of water-waves, because they are so much easier to visualise; but they are only meant for an illustration by which we aim at describing the behaviour of light. Let us suppose that instead of making two small gaps in the sea-wall, we have a single wide gap, as in Fig. 6. The principle of superposition can be applied to show what will happen, and in so doing we are using what is called Huygens' construction. We regard the wide gap as a lot of small gaps touching one another. Each of the small gaps gives out a wavelet with a set of circular crests emerging from it, and the effect at any point is the sum of all these circles. In the immediate neighbourhood of the gap there is a very confused state of affairs, which we need not consider, but farther off the pattern of the crests tidies itself up. Each separate wavelet is very small, and the only effects we can see will be those due to the co-operation of a large number of them. The principal place where a lot of the wavelets are in phase together is in the direction straight forward from the gap, as may be seen from the way all the circles touch one another there. So our construction shows, as we should expect, that when the waves come to a wide

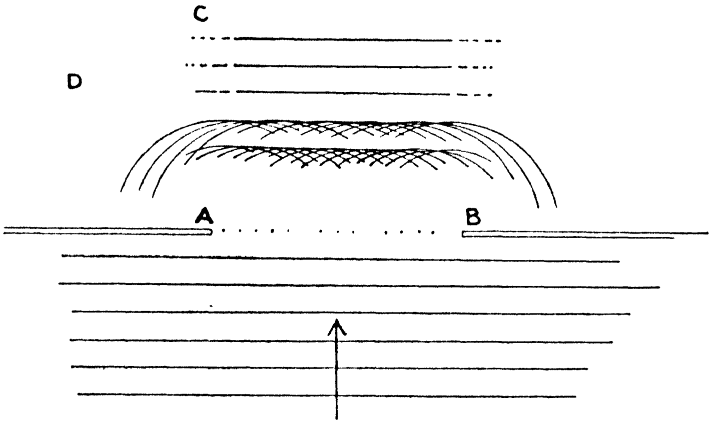


FIG. 6.—DIFFRACTION AT WIDE GAP

Circles are drawn with centres at points of AB to represent the crests of the wavelets in Huygens' construction. In the region opposite the gap the circles touch one another and reconstruct the straight crest. Towards D the phases of the wavelets are spread round a complete cycle, and so they destroy one another by interference. To simplify the figure only a few of the circles have been drawn.

gap they travel on through it in a straight line, just as though there were no wall. Towards the line of the edges *AC* this is not so, because the wavelets that ought to have come from beyond the edge of the gap are missing. Consequently there is a confused region along the line *AC* where the motion is complicated. Well away from the line of the gap, in the region *D*, we can see that the wavelets from all the parts of the gap have their phases spread all round the circle, and so they will cut out together. We can describe the whole process in the language of light by saying that the sea-wall casts a shadow; opposite the wall there is darkness, that is, no motion; opposite the gap there is light, that is, the full undisturbed wave-motion. But whereas we are apt to think of the edge of a shadow as sharp, we see that here it is not so; there is a confused region round the line of *AC* which passes gradually from the light to the shadow.

The same process is verified for light, though on account of the exceedingly small wave-length it is on a very small scale and hard to observe. The shadow of a straight edge is not sharp, and the intensity of the light in its neighbourhood is shown in Fig. 7. This is drawn for light of a single colour, and it is interesting to see that just outside the shadow there is a place where there is actually more light than in the region of full illumination. Ordinary white light consists of all the colours superposed, and as the blue has shorter wave-length than the red, the blue maximum will be a little closer in and so the edge of a shadow will look blue. The general phenomenon that light does not form sharp shadows was discovered long before Young by Grimaldi in 1665, and called by him *diffraction.* It was explained first by the great Fresnel, using the conceptions that I have been describing.

The phenomena of diffraction play a fundamental part in the new theories of matter, so we may usefully recapitulate them in the form of the question—what do we mean by a ray of light? Suppose that we make a hole in a shutter so as to admit sunlight. We then get what looks like a ray of light (which can be seen because it illuminates the dust in the air). On rough examination it appears to have fairly sharp edges, but we soon see they are not so, and find that this is because the sun has a finite size, so that the rays from the different parts of its surface are going in different directions and are spreading out from the shutter. This effect is not what we want, and we can eliminate it by taking a second shutter with a hole in it behind the other so that we only get light from a small part of the sun's surface. We now have a beam of nearly parallel light of breadth corresponding to the size of the hole in the shutters. We naturally think of this as composed of rays of light, and we set to work to isolate a single ray from the others. To do this we diminish the size of the holes in the shutters and get a narrower beam, and we can say that we have cut out

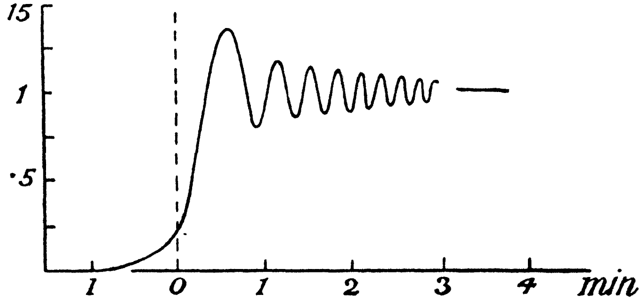


FIG. 7.—EDGE OF SHADOW

The geometrical position of the shadow is indicated by the broken line. The curve shows how the intensity of the light fluctuates in the neighbourhood of the edge. The size of the pattern is shown by the millimetre scale, on the supposition that the light is yellow and the obscuring object one metre away.

some of the rays and so have got part way towards isolating the single one. We can easily go on in this way for some time, but when we get down to a certain very small size we are surprised to find that a further diminution of the holes produces the opposite effect; instead of making the beam narrower we find that it begins to spread into a cone, and as the hole gets still narrower the cone gets wider and wider, until for an exceedingly small hole the light goes equally into all directions. In the effort to isolate a single ray we have lost it. The reason is seen by our explanation of how waves behave. There was always a confused region, the region of diffraction, at the edge of the beam, but as long as the beam was broad it could be overlooked because it was so much narrower than the beam itself. But as we narrow the beam this region of diffraction becomes relatively more important and for a very narrow beam may predominate over the beam itself. It is no use going further, for if we narrow the shutter more, the diffraction region gains in strength and spreads farther and farther out. As to when diffraction begins to become important, the determining condition is the wave-length of the light. As long as the hole is much larger than the wave-length we can speak of rays of light, but the edges of the beam become vague when the hole is a few wave-lengths broad. If the hole is two wave-lengths broad the light will spread out roughly into a cone of 30°, and if it is less than half a wave-length, it will be nearly uniform over all directions.

We have so far only described the behaviour of pure harmonic waves spreading in two or three dimensions, but must now consider more general waves which are not of pure harmonic type. This is a more complicated question, but fortunately it will suffice to consider the case of what we call a one-dimensional wave only; that is to say, we may think of the vibrations of a wire, or of sea-waves with perfectly straight crests at right angles to which they move. For harmonic waves the wave-velocity is determined by the wave-length, but we have not considered how. The relation between the two quantities depends on the medium, as was mentioned above. The simplest possible case is when the wave-velocity is the same for all wave-lengths, and this we may consider first. It is true for light in a vacuum, and for sound, and also for the transverse vibrations of a flexible wire. We may think of the last case, as it is easiest to visualise.

The principle of superposition is the guide which shows us what will happen. The wire can carry simultaneously two or more harmonic waves, each of which is propagated just as though the others were not there. Now in our special case the wave-velocity is the same for both, and so if two of their crests are together at the start, they will stay together later. The same is true if further harmonic waves are superposed; whatever the deformed shape of the wire at one time, this shape will travel along the wire unchanged in form. Now any shape of the wire can be regarded as the superposition of harmonic waves of all wave-lengths, and so we can conclude that if a wave is started, by which the wire is deformed in any way whatever, it will travel along at a uniform speed without changing shape as it goes. Though there is no transfer of matter along the wire, there is a transfer of shape. This allows us to endow the wave with a character of individuality as something perfectly real, because we can identify each part of it. Such a wave is just as individual as the things that we ordinarily regard as particles; for the essence of individuality is that one could mark the object, so that it should always be recognised again. When we want to identify a particle among other similar ones (as for instance in billiards), we do so by marking it in some way by which it can be distinguished from the rest; in the same way here we can label one particular piece of the wave, say by giving a small extra kink to the wire, with the certainty that this kink will travel at just the same speed as the rest of the wave, and so continue to serve as an identification mark.

Just the same result is true for a wave of sound, [illegible] light, as long as we are dealing with a wave of one [illegible]mension, in which the phase is the same all over each [illegible] a set of parallel planes. But in two or three dimensions [illegible] when light passes through a hole in a shutter, we have seen that this is not so, because diffraction produces a complicated change of form at the edge of the shadow. It would not be possible to trace the individuality of a particular crest by marking it with an extra kink, because this kink would get muddled up with the other crests by the diffraction at the gap. The same failure occurs even in one dimension, when the waves are in a medium for which the wave-velocity depends on the wave-length, and this case is important for our theory. A very good example of the point is furnished by the question sometimes asked—how long does it take for a signal on the submarine cable to cross the Atlantic? There is no definite answer to the question, because the sound heard at the other end is a different sound from that which started. If the signal is a sharp click, at the farther end a faint sound will begin a little later, grow to a maximum, and then fade away again; we cannot say which part of this noise is really the signal, and so we cannot say how fast it has gone. Nevertheless we can attribute an inferior kind of individuality to the whole signal, and this we will now consider.

Water-waves are one of the types for which wave-velocity depends on wave-length, and so we will consider a train of water-waves limited in length, large in the centre, and diminishing towards both forward and backward ends. As we watch it progressing we shall see the crests moving forward at the speed of the wave-velocity. But we shall also notice another thing, that as they go they change in height. The central wave-crest was at first highest, but as it goes it shrinks in height and the one behind it becomes higher. The same is true all along the train; the foremost crest which was already very small becomes too small to see, and the hindmost small crest becomes bigger, while another new one appears behind it. The general effect is that the *group* of waves goes slower than the individual waves—or we say the *group-velocity* is less than the *wave-velocity*. Anyone who has watched the bow-waves made by a motor-launch will have noticed that they seem to be moving quite fast, and yet that they take a long time to travel across the water. In watching them we watch a single crest; this fades out as it goes, and it is the group-velocity that determines the rate at which the general disturbance travels.

The phenomenon of groups of waves depends only on the fact that in many media harmonic waves travel at a speed depending on the wave-length. In water long waves go faster than short, and the group-velocity is less than the wave-velocity—it is actually half. For the vibration of an elastic bar the opposite holds, short waves go faster than long, and so the group-velocity is greater, actually it happens to be double. Fig. 8 is a drawing of successive positions of a wave going along a bar, the initial form being chosen to make the calculations easy. We can see how the small trough in front is growing at the expense of the main crest. The centre of the patch is moving forward twice as fast as the crests of the waves. The case of water-waves and of a bar have the specially simple relation of half and double for the group- and wave-velocities, but there may be in a suitable medium any relation between them. There is even no reason why they should be in the same direction; it is quite possible to have a set of crests moving to the right, but fading out as they go so fast that the centre of the wave region is continually shifting to the left.

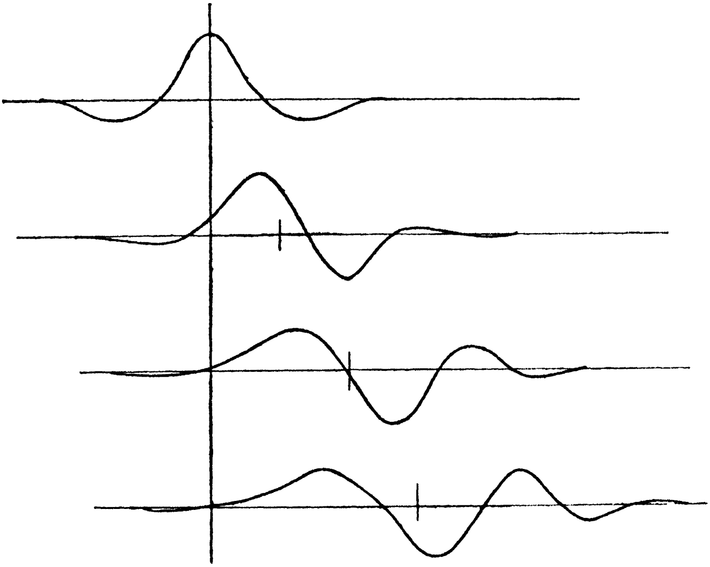


FIG. 8.—WAVE ON STIFF BAR

Successive shapes assumed by an elastic bar carrying a wave. The short cross-lines show the position of the centre of disturbance, which moves twice as fast as the individual crests. The length of the disturbed region increases as the wave travels.

We can see in Fig. 8 that in addition to the motion of the waves there is a certain amount of spreading, so that if we measure the length of the disturbed region at any time we find that it grows progressively. The rate of growth depends on the length of the region itself, and if the region is at first very short it spreads rapidly; but if we have a long train of nearly harmonic waves, then we find that the spreading is much slower. The quantity that determines the rate of spreading is the number of approximately harmonic waves in the train; if there are many there is little spreading, but if few it spreads rapidly. In Fig. 8 there was only a little more than a single wave, and we can see the spreading almost immediately.

When we discussed diffraction we saw that we were entitled to think about rays of light, provided that we had no obstacles as small as the wave-length. In that case it is legitimate to forget about the confused region at the edge of the beam, and to attribute a sort of individuality to the light as it travels, tracing out definite paths for the rays. We are now in a position to do the same for the fore-and-aft direction with the help of the idea of group-velocity. The case of light is really unduly simple, because for it the group-velocity is the same as the wave-velocity, but we will nevertheless use the language appropriate to optics, premising that the results are true for any waves.[1] Suppose that we open a shutter for a very short time, so that a flash of light passes through. This will consist of a group of waves of limited length, and will therefore advance with the group-velocity. As it goes it spreads, but only slowly if the group includes a considerable number of crests. In that case it will be possible to attribute individuality to the group, in the same sense as we attributed it to the rays of light passing a large hole in a shutter, and we can think of rays of light travelling at a definite speed. But if the group only includes one or two crests, the spreading will be very rapid, which will make it impossible to say how fast the ray is going, as in our example of the submarine cable. This is just like the previous case, where we saw that if the gap was only a wave-length broad, the wave would spread out in all directions, and that there, too, it was meaningless to speak of rays of light.

---

[1] For light in glass the group-velocity is not equal to the wave-velocity, and the present description would apply.

We may summarise this description of waves by saying that their behaviour depends entirely on the scale of their wave-length in comparison to the various obstacles they encounter. If the holes, moving shutters, etc., are all devised on a coarse scale, so that they always admit large groups of waves, then the groups can be regarded as having a kind of individuality—a given patch of disturbance can be traced through successive positions. Such a patch goes on in a straight line at a definite speed, just as would a particle. To exhibit the wave-characters it is essential to have obstacles of a scale nearly as small as the wave-length. It was the fact that they had no obstacles small enough that permitted the physicists of the eighteenth century to think of light as composed of particles. It was only a few years ago that obstacles were devised so small that they could exhibit the wave-characters of ordinary matter.

CHAPTER III

THE DIFFRACTION OF MATTER

We must now examine quite a different aspect of nature, and one that is much more familiar. What is the ordinary motion of an ordinary body? The main principle is contained in Newton's First Law, which says that every body remains at rest or in uniform motion in a straight line, unless acted on by a force. If this is examined critically it is mostly a matter of definition. For example, how are we to know that it is not being acted on by a force? Because it is moving uniformly. Or again, how can we verify that its speed is uniform? By a clock. How do we know that the clock keeps time? By the laws of dynamics. How do we know the laws of dynamics? Why, from Newton's First Law. But we may, I think, omit these considerations here, and whether we regard the law as a definition, or as an observation made by reference to intuitive ideas of time and space, we can regard the law as the starting-point of dynamics.

The most interesting thing about this law is its mention of a straight line. We ordinarily think of a straight line as the shortest distance between two points, and if we are required to construct one, we take a thread and pull it tight. So we have now two ways, apparently radically different, of constructing a straight line. We have also seen a third process which yields a straight line, and that is a patch of approximately harmonic waves in a medium; as we saw, this also moves at a uniform speed in a straight line. We thus have three processes in nature, apparently radically different, but all accomplishing the same task, that of drawing straight lines. When different processes yield the same result, it is necessary to inquire whether they are really different, or whether a single principle may not underlie them all. It was largely this inquiry that led to the elucidation of the principles of the new mechanics.

If dynamics were only concerned with rectilinear motions it would be a delightfully easy subject. But of course the interesting motions are those in which forces are pulling the body out of the straight line, for example, the ordinary motion of a projectile. The usual description of this is to say that gravity is pulling the particle downwards with a constant acceleration, and that gives a complete account of the process, but we seem to have lost the analogy with the stretched string. Now this analogy is so good that we naturally try to keep it; to understand how this can be done we may first take a rather different case. Suppose that I am considering the motion of a particle not in free space, but on a smooth curved surface. We cannot then regard its motion as a straight line, since it has to keep on the surface which is not flat. Nevertheless it can be shown that it will in fact describe a path which is the same as that made when a string, confined to the surface, is pulled tight. This is called a *geodesic* line, and is the next best thing to a straight line that exists on the surface. So here again we can define the path by a *minimum* property; it goes between two given points by the shortest route. We can do the same sort of thing for our projectile, though there is no such simple geometrical description for it. There is a certain quality of any motion which a particle could perform, and for the one which is actually performed it is a minimum. This principle is called the Principle of Least Action.

The Principle of Least Action has been discovered a number of times in different forms, and it is interesting to consider them, for they show what various aspects it may take. Its first description, due to Hero of Alexandria, must rank as one of the earliest scientific laws discovered. In describing how light is reflected by a mirror, Hero pointed out that it pursues that path from source to mirror, and then to eye, which takes the shortest possible time. Its next appearance was also in optics, and was due to Fermat early in the seventeenth century. He was discussing the passage of light through transparent media, and showed that the track of the light, from a point in air, say, to a point in water, can be found by supposing that it takes the shortest possible time, provided that allowance is made for a slower speed of the light in water than in air. The principle next appears in dynamics early in the eighteenth century, and is due to Maupertuis, who called it the Principle of Minimum Work. He gave certain rules for calculating the "work," for any path that the particle might pursue, and showed that it would actually follow the one for which this was least. His guiding idea was a sort of metaphysical principle (which he attempted to justify on theological grounds), that Nature, like humanity, was lazy and took care to do as little work as possible. He succeeded in showing that his principle would lead to all the results of dynamics. Alternative forms of the principle were later given by Lagrange and by Young; but we may pass on to Hamilton, who finally tidied up all the previous work, showed how the various formulations were related, and in fact finished the subject off.

We had better describe the rules of operation a little more definitely. Suppose that we want to find the path of a projectile under the influence of gravity between the points *A* and *B* of Fig. 9. In the first place it is obvious that there are a great many natural paths possible, according to whether we shoot a bullet with a nearly flat trajectory or lob a ball in the air. In order to be able to get a definite answer we must therefore restrict our conditions, and shall do so by saying that the projectile is to go from *A* to *B* in some definite time. Choose any path whatever, say *ADB*, and assign a speed at every point on it, subject only to the condition that if the projectile wastes too much time over the first part it may have to do the rest in a hurry so as to get to *B* at the proper time. Now construct a certain quantity depending on the speed and position (actually it is the *difference* between the kinetic and potential energies at each point of the path). The values of this quantity all along the path are then to be averaged over the time from *A* to *B*. The result is the action for that particular path. Of all possible paths, such as *ACB*, *ADB*, *AEB*, and all possible variations in the speed along them, the one which gives the least value for the action is the parabola *APB* which the projectile actually describes.

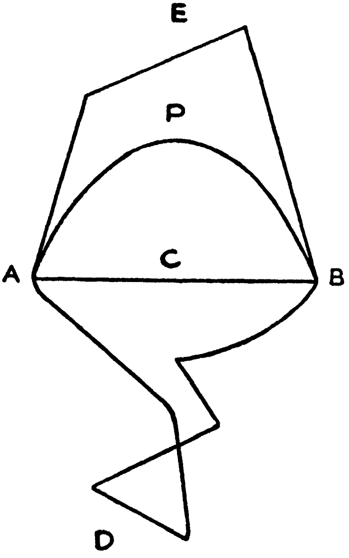


FIG. 9.—PRINCIPLE OF LEAST ACTION

The action is evaluated for every path from *A* to *B*, and found to be least for the parabola *APB*.

We have only outlined a particularly simple example, but Hamilton applied the same process to any system, however complicated, and so made it into a general and quite fundamental dynamical principle. It is so universal that from his time onwards it has been regarded as the ultimate principle of Nature. And yet there is something unsatisfactory about it, even if we accept Maupertuis' idea that Nature is inherently lazy. For if that were so, then we should expect that Nature would only succeed in economising labour after the same fashion as does humanity. Now when *we* want to economise, the only way in which we can do so is by making a number of different trials, and seeing which is cheapest. Let us imagine ourselves in the position of the projectile and consider how *we* should learn to get from *A* to *B* with the expenditure of the least possible action. Our equipment must include a stop-watch, so as to be sure that we get to *B* at the right time. Then we need a speedometer to measure our kinetic energy, and an altimeter for our potential energy. We also need a large amount of paper to work out the result, unless perhaps we can devise a calculating machine to do the work automatically. With this equipment we go from *A* to *B* along some path and evaluate the total action. We next try a number of other routes and see which of them is cheapest in action. Next we must select a number of routes near this cheapest one to see if any further improvement is possible; and thus after a large number of trials we should arrive at the result that there is a certain path, a parabola, which is more economical in action than any other. This is what *we* should have to do in order to obey the Principle of Least Action, whereas Nature is able to do it at once. In fact, we are forced to admit that even if Nature *acts* in a lazy way it must *think* in a highly intelligent and by no means lazy way, before its thought can take practical effect.

The power and convenience of Principle of Least Action had the consequence that during most of the nineteenth century this philosophical difficulty was forgotten, but as a matter of fact the difficulty could all the time have been met by consulting the original works of Hamilton. These had been written under the impetus given by the great developments of the wave-theory of light. Let us consider the simplest case of least action, the motion of a particle in a straight line. Here we could not only identify the motion with the shortest distance between two points, but also with the motion of a group of waves in a medium, which also travels uniformly in a straight line. Now in discussing diffraction we saw that when the group came to a barrier a good deal wider than its own wave-length it could pass through more or less undisturbed, but that this transmitted wave could alternatively be regarded as built up out of a large number of wavelets given out by each point of the gap.[1] We may think of these wavelets as a set of scouts thrown out by the group, exploring to find the way it should go, and at once the Principle of Least Action becomes natural. The scouts are all the time making trials to find how the action can be made least; the wavelets cancel out everywhere except at the places where the action will be least, but co-operate along the path of the group. We see that Nature practises economy by trial and error in the same way that we do ourselves.

Hamilton set himself to extend the idea to cases where the particle does not move in a straight line because it is being influenced by forces, but he restricted the problem by only dealing with waves of such short wave-length that there was no need to trouble about the confused region at the edge of his patch of waves. In this case, as we saw, it was legitimate to speak of rays of light, and he worked out a general method of treating rays of light (still largely used by the practical designers of optical instruments), which was immediately applicable to the motions of particles of matter. So Hamilton's theory, which started as optics, changed over and became a theory of general dynamics, and it was for a long time almost forgotten how it had originated.

---

[1] See Fig. 6, p. 43.

We have so far only shown how the Principle of Least Action agrees with wave theory in the case of uniform motion in a straight line. If that were the only example of agreement the result would be too special to be of any use, but it is easy to extend it to more general cases. In considering waves we have only discussed their propagation in a uniform medium, and must now see how they behave in variable media. The most familiar example is the passage of light from air to water. At the surface there is *refraction*, a sharp bend in the direction in which the waves face. But we need to consider a more general case than this, in which the qualities of the medium change continuously from point to point, and an admirable example can be drawn from the behaviour of water-waves.

The rate at which waves travel on water depends on the depth of the water. We avoided this complication before by only considering deep water, but we will now take the depth into account. As a wave advances into shallow water its speed slows down, and consequently the crests tend to crowd together. Let us now imagine that we have a group of nearly harmonic waves out at sea, running in any direction, and examine what happens to them as they approach a shelving beach. In Fig. 10 we make Huygens' construction by drawing a small circle round each point of a crest to represent the position which would be occupied say a second later by the circular wavelet sent out from that part of the crest. At the shoreward end the circles will be smaller because the wave goes slower there, and so we see that the waves wheel round so as to face more towards the shore. This explains why it is that, no matter what the direction of the wind, waves always come in straight towards the beach. Furthermore, if we have an isolated patch of disturbance on the water forming a group of waves, we shall find that when it approaches shore it will wheel round and move more directly shorewards. We may describe what happens by saying that the group behaves as if it were acted on by a force attracting it towards the shore.

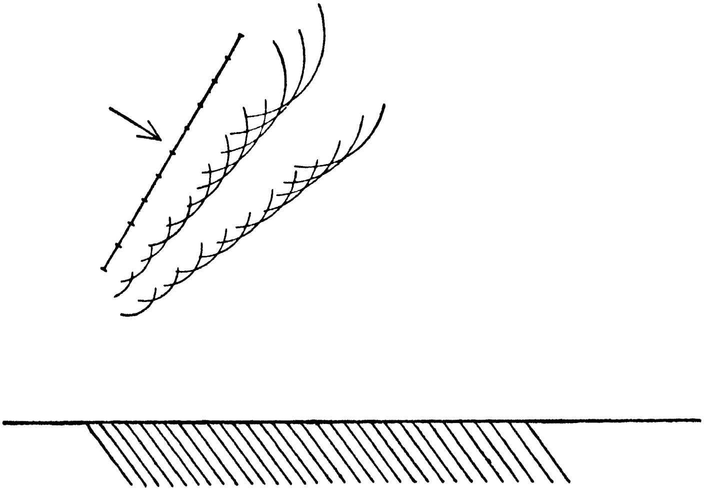


FIG. 10.—WAVES ADVANCING TOWARDS SHORE

In Huygens' construction the circles at the shoreward end are smaller than those at the seaward end, and the waves wheel round towards the shore.

This example will serve to show how least action works when there are forces acting on a particle. The particle is to be likened to a wave moving through a medium which possesses properties such that the wave-velocity varies from one point to another; in optical parlance the refractive index is variable. A case may be cited where this actually happens in a striking way.

The air over the sea under certain circumstances gets arranged in layers of very unequal temperature, and it may happen that the wave-velocity in the upper air is a good deal higher than that at sea level. Under these conditions it is possible that light can get more quickly from distant land to a ship by going up and coming down again, instead of going along horizontally. The rays of light will therefore pursue the upper course, and the sailor will see a mirage. As illustrated in Fig. 11, the land will be seen up in the sky in the direction from which the rays are approaching him. Hamilton's description of the path of a projectile would be very like this; the curved path would be attributed to the inequality in the wave-velocity, set up by the gravitational force of the earth. This force is always there to bend the path of a projectile downwards, but it is only the chance of a peculiar distribution of temperature in the air that bends the light-waves and produces the mirage.

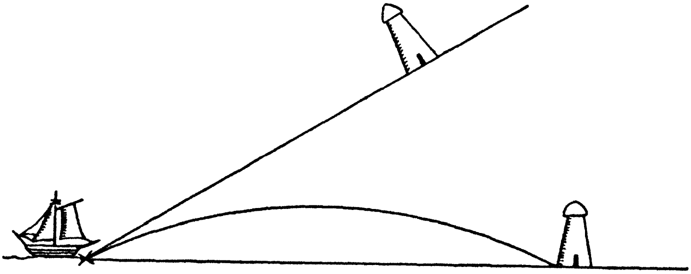


FIG. 11.—MIRAGE

Light going from the lighthouse to the ship travels quicker by the curved path than horizontally, and so the lighthouse appears lifted into the air.

As I have said, Hamilton limited himself to cases where the wave-length was so short that diffraction could be disregarded, and his principle was so useful that for a long time no one looked further into the matter. The Principle of Least Action served as the basis of all dynamical ideas all through the nineteenth century and for the first quarter of the twentieth, though by that time it had become clear that something was wrong. The behaviour of atoms would not conform strictly to the deductions from the Principle of Least Action. This period was a very curious one in the history of physics. A great body of doctrine, the Quantum Theory, came into existence, which was successful in explaining so many phenomena that no one could doubt its validity, and yet all the time it seemed to rest on inconsistencies and logical contradictions. There were no foundations and yet an enormous and obviously well-built superstructure. I shall not, however, describe any of the details of what is now called the Old Quantum Theory (which only died about five years ago), as most of its inconsistencies have been entirely cleared up.

Though many had a hand in the solution, the key idea, which makes common sense and not merely mathematical sense of the theory, was due to the genius of Louis de Broglie. He had felt the philosophical difficulty about least action which I have described, and, after the manner of Hamilton, proceeded to work out analogies between particles and waves, but with the great advantage over his predecessor of having a vastly greater field of physical experiment to draw from. His analogies were only given a mathematical form later, but when this had been done it became clear that the principles of dynamics go wrong for atoms and electrons on occasions when the obstacles which they encounter are very small, and would be put right by allowing for the diffraction of waves. The really terrible difficulties of the Old Quantum Theory could all have been avoided by a fearless application of the idea that the Principle of Least Action is a wave principle. In saying this I do not mean to imply that physicists were stupid in not perceiving it earlier; for, as we shall see, de Broglie's work implies a very complete revolution in our ideas of the nature of matter. If his suggestion had been made even only a few years earlier it would hardly have been acceptable, for a great many of the properties of matter would have appeared very definitely to fit in with the old ideas and to be incapable of explanation with the new. With the reconciliation of the two aspects we shall be concerned in the next chapter; for the present we must consider what direct evidence there is that atoms and electrons have the characteristics of waves.

We have seen that a group of waves behaves like a particle as long as the slits, etc., that it passes are on a scale many wave-lengths broad. To evoke the wave-character, obstacles must be used so fine in structure that a perceptible amount of diffraction will occur. In the eighteenth century no obstacles were constructed even fine enough to bring about the diffraction of light, and it was therefore legitimate to think of corpuscles of light, and it is only very recently that the diffraction of matter has been observed. The difficulty in both cases was the same in principle, that the very finest things which we can make with our fingers are still far too coarse in scale. We must therefore consider what methods have been used to produce instruments which will yield an observable amount of diffraction. We shall consider the simpler case of light first.

The standard instrument for studying the diffraction of light is the "diffraction grating." The earliest gratings were made by Fraunhofer at the beginning of the nineteenth century, by mounting two fine screws parallel to one another and winding a wire to and fro between their successive threads. At the present time they are made by ruling many thousand parallel lines on glass with a diamond, and such gratings are the indispensable instruments of the science of spectroscopy. To consider how they work, we may take a simplified model. Imagine that we have an opaque screen with a considerable number of fine slits cut in it, all parallel to one another and equally

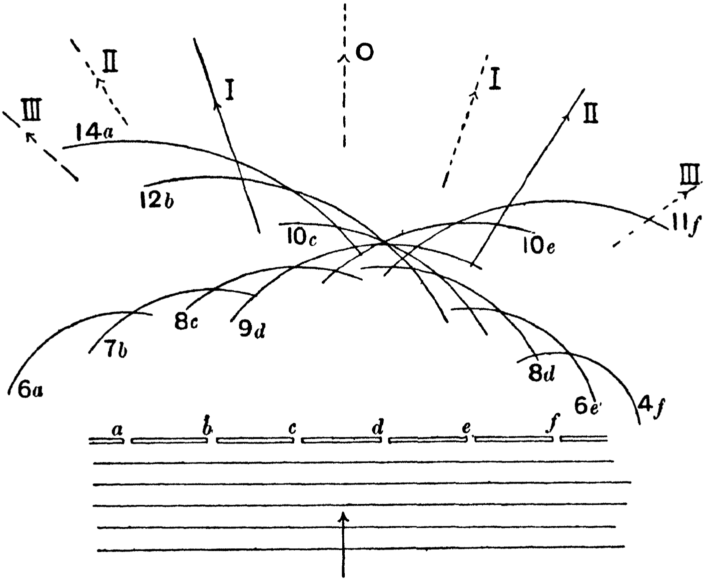


FIG. 12.—DIFFRACTION GRATING

The wave-crests are shown falling on the grating *ab* . . . *f*. From each hole circular waves emerge, but only a few are shown. Thus 4*f* and 11*f* are parts of the fourth and eleventh crests emerging from *f*. The circles 11*f*, 10*e*, 9*d*, etc. reinforce one another and give rise to the *first order spectrum* I, while 4*f*, 6*e*, 8*d*, etc. give the *second order spectrum* II. The broken arrows indicate the directions of other orders of spectrum which could be exhibited by drawing more of the circles.

spaced. Waves of light of a single wave-length fall on one side of it; let us consider what happens on the other side. From each slit circular wavelets will emerge, and we draw a number of circles to represent the positions of their crests at some later moment. Fig. 12 is rather like Fig 6, which we had when we considered the diffraction of a wide gap, but differs from it in one essential respect.

With the wide gap the circles we drew had their centres close together all along the gap, here they are separated by the opaque parts of the screen. We search for places where the various wavelets reinforce one another, and find as before that they do so in the forward direction, but we can see that there are now other directions as well. The smallest circle round *a* is touched by the second smallest round *b*, and by the third smallest round *c*, and so on, so that in this direction the wavelets from all the slits will reconstruct a new plane wave going at right angles to these circles. This wave is called a *lateral spectrum* of the grating; its position may be described by the rule that the phases of waves from any two adjacent slits differ by a single cycle. There is another direction more oblique where the phases of waves from adjacent slits differ by two cycles, and this wave is therefore called a *second order* lateral spectrum. There are, of course, also lateral spectra on the other side of the direct line.

The general effect of a grating is thus to change a single incident wave of light into a set of waves going in a number of different directions. The directions depend on the ratio of the wave-length to the *grating constant*, which is the distance between adjacent slits. If the grating constant is large, the lateral spectra are close together; they can easily be seen by looking at a distant light through a sheet of fine muslin, but the pattern is more complicated than I have described, because there are two sets of threads at right angles instead of a mere line of slits, and because the light is white instead of monochromatic. If the grating constant is made progressively smaller, the spectra move out to broader angles. It is easy to see from the construction that if the constant is for example twice the wave-length, then the second order spectrum will be in the direction which grazes the plane of the screen, and there will be no orders at all of spectra higher than the second. On still further diminishing the distance between the slits the second order disappears and the first order moves out to still wider angles, and when the constant is equal to the wave-length, it in turn grazes the screen. For still smaller grating constants there are no lateral spectra at all; the light goes right through in the direct line only. This illustrates an important principle of optics, that to see the features of anything, it must be illuminated with light of a wave-length shorter than the details to be detected. From these general considerations it appears that if we want to observe the diffraction of waves easily, there must be a fairly precise relationship between the wave-length and the grating constant. A grating finer than the wave-length is no use at all, and one that is very coarse, say with constant a hundred times the wave-length, will give lateral spectra very close to the direct beam, which will be hard to see separate from it.

There is no essential difference theoretically between the effect of a grating and that of the diffraction bands produced by two holes which we discussed in the last chapter,[1] since both are derived from the same principle, the principle of superposition. But there is an enormous practical advantage in a grating. The bands from the two holes are very faint and close together, whereas the thousands of slits of a grating all co-operate in producing the lateral spectra, and may make them very brilliant and easy to observe. On this account all practical work on diffraction makes use of gratings. On the other hand, in considering theoretical matters, it is easier to think of the simpler arrangement with only two holes, which produce a set of diffraction bands crossing the illuminated screen. So we must remember that the effect observed with the grating directly implies that the corresponding effect would be observed with two holes if we could make the appropriate experiment. We shall be discussing various diffraction effects experimentally observed with gratings, but it will facilitate their interpretation to translate them into the effects which would have been seen for two holes, even though the actual experiment with two holes may be quite impossible for practical reasons.

---

[1] See Fig. 5, p. 40.

Early in the nineteenth century gratings were constructed which would resolve ordinary light and show its wave-character and measure its wave-length, and this gave rise to the great science of spectroscopy. I would like to enlarge on this science, because it has had more effect than any other on the development of the quantum theory, but it would take long to do so, and the principles of the quantum theory can now be explained without discussion of the rather intricate details of spectra. We may be content with the more general property of gratings, the fact that when they give rise to lateral spectra, they prove that we are dealing with waves. Now in 1895 Röntgen made the great discovery of X-rays, which inaugurated the new era in physics. For a long time their nature was rather mysterious. They seemed to go perfectly straight like particles and to have other characteristics of particles to which we shall come later. On the other hand they travelled at the speed of light, and general ideas pointed to their being some sort of light. If they were really waves like light, they ought to show diffraction, but their wave-length must certainly be much shorter than that of light, and for a long time no one could devise a grating fine enough to test the matter. Finally in 1912 it occurred to Laue that Nature had provided us with gratings thousands of times finer than any artificial grating, and by these means he showed the diffraction of X-rays and established their identity with light.

These natural gratings are the crystals which compose most solid bodies. In a solid body the atoms are packed as close together as possible, and this forces them to fall into a pattern. The effect may be illustrated by taking a lot of marbles of equal size; if they are packed tight together they will all fall into rows. In a crystal we get the atoms falling into rows and the rows into sheets, all perfectly arranged at regular distances apart, as in Fig. 13. The distances between the sheets are determined by the sizes of the constituent closely packed atoms; they are about two or three hundred millionths of a centimetre apart. Now consider what happens when a wave falls obliquely on a crystal. Take first the sheet at the surface. In Fig. 14 *P* is a source of light or X-rays and *LMN* are the atoms of the sheet. Each atom scatters a wave in phase with the incident light, but *M* is farther from *P* than *L* by the amount *MH*, and so the phase of the wave scattered from *M* will be that amount behind the one scattered by *L*. We therefore draw round *L* a circle of some radius *LS* to represent the later position of one of the crests at some instant of time, but round *M* we must draw a circle of radius *MT* less by the amount *MH*. Round *N* we draw a circle of radius smaller than *LS* by the amount *NK*. We can then see that these crests will

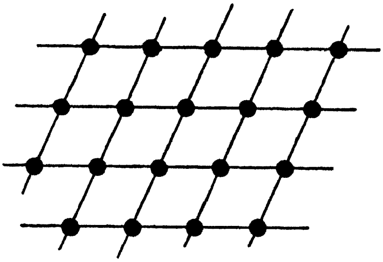


FIG. 13.—LATTICE OF ATOMS

co-operate along a front *STU*. This is the explanation of the ordinary reflection of light; it would be true for an ordinary mirror, and so far it tells us nothing new. The easiest way of thinking of the geometry of reflection is to make use of the image point *I*; if the light emerges from a source at *P*, and is scattered by the plane *LMN*, the scattered light behaves just as if it had come from the place *I*, the image of *P* in the plane.

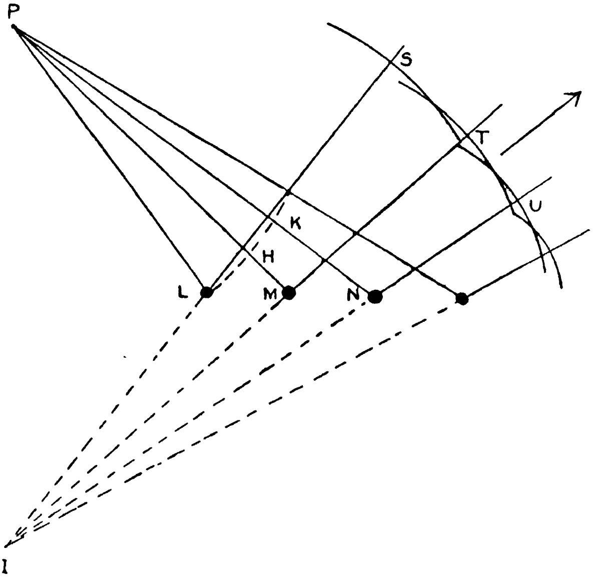


FIG. 14.—SCATTERING OF X-RAYS BY A SHEET OF ATOMS

The light at *M* has phase *MH* behind that at *L*. When the crest of the wave scattered by *L* has radius *LS*, that from *M* will have radius *MT* less than *LS* by amount *MH*. The total scattered wave has front along *STU*, which is as though it came from the image point *I*.

Now take a whole crystal. Instead of a single plane of atoms *a* at the surface, we shall have a regular succession of parallel planes *a*, *b*, *c*, . . ., as shown in Fig. 15. Moreover, each plane only reflects a very little light, so that *b*, *c* . . . are hardly screened from the incident light by the layers above. The light reflected by *a* behaves as if it came from the image point *A*, but there will also be a reflected wave from *b*, behaving as if it came from *B*, another for *c* from *C*, and so on. The resulting reflected wave is the joint effect of all these waves superposed, but in superposing them we must remember that they have definite phase relations with one another. The result will

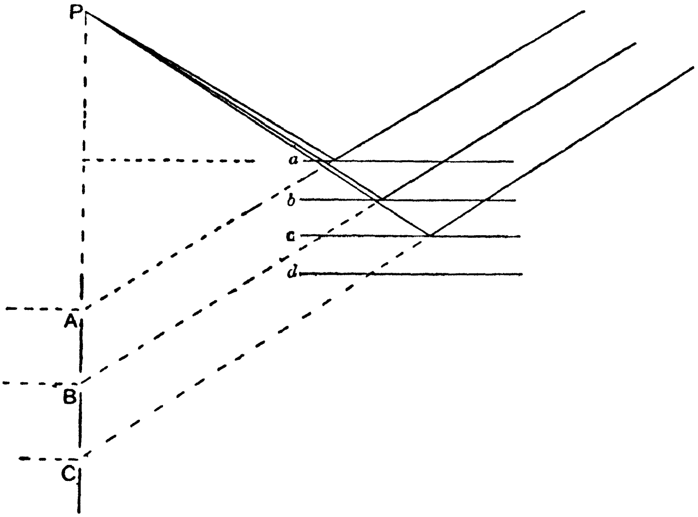


FIG. 15.—SCATTERING OF X-RAYS BY A CRYSTAL

The light scattered from the layer *a* behaves as though it came from *A*, that from *b* as though it came from *B*, and so on. The resulting wave is the superposition of these, and may be regarded as a lateral spectrum of the grating *ABC*, such as was shown in Fig. 12.

be exactly as though we had light falling on a grating containing holes at *A*, *B*, *C* . . ., and the reflection from the crystal will consist in the lateral spectra of this grating. If the crystal is set suitably, the wave from the plane *b* will be exactly one wave-length behind that from *a*, that from *c* will be two wave-lengths behind, and so on, and they will all reinforce one another. For other settings their phases will not be in agreement and the waves will cancel out. Hence the general effect is that a crystal can only reflect X-rays of a definite wave-length if it is set at one of certain definite angles, which correspond to the lateral spectra of first, second . . . order, which would be formed by the grating *A*, *B*, *C* . . . .

As soon as the experiment was tried, it was found that there were very strong reflections at certain definite angles, and these corresponded to the various orders of spectra in the manner I have described. This result was the starting-point of two wholly new branches of physics, for in the first place it made it possible to measure the wave-lengths given out by X-rays from different elements, and in the second it furnished a method of finding out how the atoms are packed in any kind of crystal; the X-rays provide a sort of indirect microscope for seeing where the atoms are. These are the practical results, but what is explicitly established by the experiments is that the X-rays are a wave-motion with a wave-length rather smaller than the distances between the sheets of atoms. That they are really light of very short wave-length depends on other considerations into which I shall not enter. The point that must be emphasised here is that they show diffraction from the crystal, and from this we can be quite certain that if it were practicable to make two tiny holes very close together, and send X-rays through them so as to fall on a screen, this screen would be crossed by alternate light and dark bands, which would correspond to the places where the two waves were respectively in the same and opposite phases.

Until very recently it never occurred to anyone to look for a similar effect among what, for want of a better name, we must call material particles. The first successes of the new quantum theory were theoretical, on the lines of applying de Broglie's wave conceptions to the electrons moving about in atoms. But it was not long before more direct experiments were devised which exhibited the diffraction of electrons which were free and in no way bound to atoms. The first success was due to Davisson and Germer. In the course of some experiments designed to find out how electrons are scattered from a sheet of nickel, they had a minor catastrophe, and their apparatus broke. The repair involved a strong heating which changed the nickel surface into a more or less perfect crystal. At the next repetition of their experiment the electrons were scattered systematically into certain definite directions, and they recognised the peculiarity as being due to diffraction by the crystal of nickel. We will not consider Davisson's work further, because not long afterwards an alternative method was invented by G. P. Thomson, which, though not essentially different, is rather simpler. To understand this, we must return to the consideration of X-rays.

We have so far considered the effect of a single crystal on the X-rays, but most matter is composed not of a single crystal, but of a large number of small, often microscopic, crystals stuck together. How will such a mixture of crystals scatter X-rays? Each of the little crystals has its planes pointing in certain directions, and for the most part they will not be set right for reflecting the rays. But among the innumerable little crystals there will be a few here and there which happen to be set just right, and each of these can reflect the rays. In reflecting them it turns them through a definite angle, and, as shown in Fig. 16, there will be formed round the main central beam a cone of rays due to the reflection from these particular crystals. If a photographic plate is put opposite, there will be formed on it, firstly, a very intense central spot of unscattered light, and then round it a circle where the cone cuts the photographic plate. But there is more than this, as we can easily see by considering the structure of the crystal in a little more detail. We have only thought of the crystal as a single set of parallel sheets, whereas it is really composed of atoms arranged in a three-dimensional lattice. Fig. 13 shows at once that we can draw other sets of parallel planes, by selecting a different set of atoms for the first sheet, and each of these sets of planes is capable of giving X-ray reflections. The distance between adjacent sheets in the new direction is not the same as in the old, and so the new set will reflect the X-rays at a different angle, and will produce a new cone of reflected rays. There are a large number of ways of drawing sheets of atoms in the crystal, some of which contain a great many atoms and so give strong reflection, and some fewer and so give a

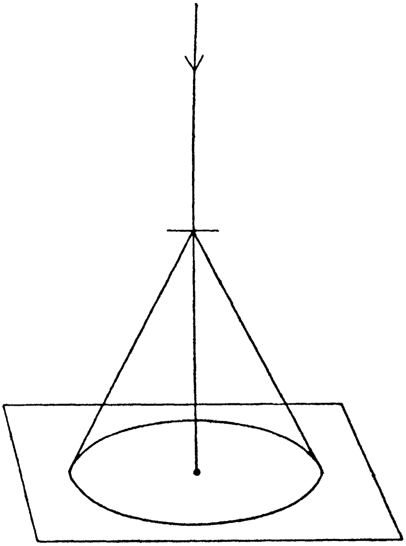


Fig. 16.—Powder Photograph

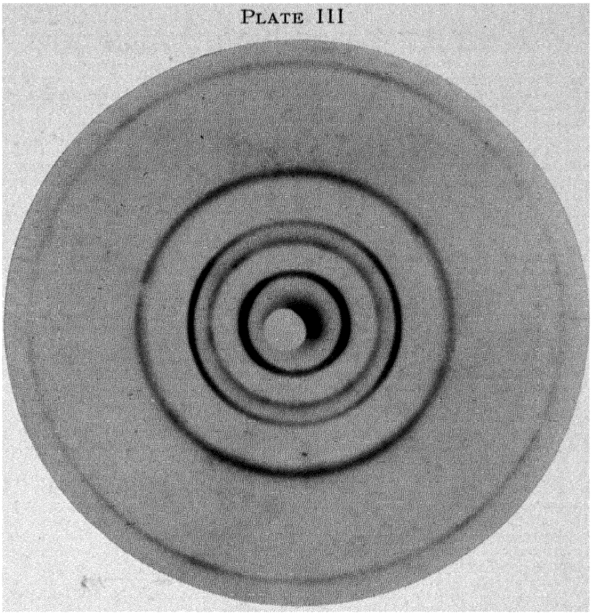
The little crystals in the powder scatter the X-rays, turning them through a definite angle, and so forming a cone round the central unscattered beam.

weaker, but each set will give a cone of X-rays, and therefore a circle on the photographic plate.

This is the principle of the "powder photograph," of which Plate III, A is an example. It is one of the most powerful ways of analysing the structure of crystals. That is not our purpose here, however; the conclusion we draw is that when we see the circles on the powder photograph, their presence proves that the rays which caused them were showing diffraction. We can confidently follow through a chain of reasoning in the following steps. The circles on the plate imply reflections at certain definite angles in the powder. The crystal structure is known, and so these angles can be verified as those corresponding to the lateral spectra of certain gratings. The lateral spectra of gratings can only be explained as due to the principle of superposition. Therefore the circles imply the principle of superposition. As a matter of convenience we then retrace the steps and construct a simplified ideal experiment, which would describe exactly the same property. If the rays were sent through two holes in a shutter, and were allowed to fall on a screen beyond it, this screen would be crossed by alternate light and dark bands.

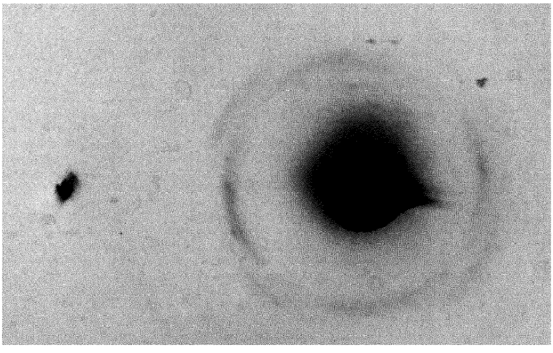
Now apply to matter the method of the powder photograph, so as to see whether we can make it exhibit the property of superposition. It is first necessary to consider what are the most favourable conditions for the experiment. Our general description of the behaviour of waves showed that a group of waves could be mistaken for a particle if all the obstacles it meets are coarse. We have hitherto always thought of matter as particles, and so, if we were mistaken, the reason must be that the obstacles we used were not fine enough; or, in other words, the wave-length must be very short. Now de Broglie's work

PLATE III



A. DIFFRACTION OF X-RAYS

A beam of X-rays passes through a powder of small crystals and falls on a photographic plate. The rings are due to the diffraction of the X-rays by the small crystals. To prevent fogging, a round hole has been cut in the plate at the place where the intensely black central spot would fall.



B. DIFFRACTION OF ELECTRONS IN A MAGNETIC FIELD

The ring system on the right was produced in the same way as those of the frontispiece, but here in the presence of a magnet. Afterwards the magnet was removed and a short exposure was made. The central spot is seen on the left. The shift of the ring system by the magnet shows that it was made by electrons and not X-rays.

indicated very exactly what the wave-length should be for atoms or electrons; we shall discuss this later. He showed that the longest wave-length would be given by the lightest particle going as slow as possible. The lightest existing particle is the electron, so we take that, but the other condition is more troublesome. It is not enough merely to submit the electron to conditions under which it is diffracted; it must be able afterwards to give some record of the fact that it has been diffracted, say by blackening a photographic plate. It can only do this if it has a rather large energy, which implies a high speed and so a short wave-length. Thus the experiment must be a compromise between long wave-length, which means easy diffraction, and short wave-length, which means easy observation. Theoretical calculations showed that such a compromise was possible, and that powder photographs could be made with electrons.

We can now come to Thomson's experiments. He took a narrow pencil of electrons at a rather high speed—they are produced by a vacuum discharge with an induction coil, and are roughly constant in speed. These pass through a fine tube so as to produce a narrow straight beam, which then falls on a plate of small crystals. Here a new trouble arises. The great virtue of X-rays is that they can penetrate easily through matter, and quite thick plates can therefore be used; but a very thin sheet of matter is sufficient to stop electrons entirely. So it is necessary to construct an extremely thin film to act as the "powder," and this is technically not an easy process. Thomson succeeded in making metal films, strong enough to hang together, but so thin as to be nearly transparent. They were probably about a thousand atoms thick, a hundred-thousandth of an inch or so. Beyond the film a photographic plate is placed, and the image produced consists in a central spot and a set of circles round it. Examples are shown in the Frontispiece. The circles are not so sharp as those from X-rays on account of the much greater experimental difficulty, but they all lie in the same places and have the same intensities. The similarity establishes that it is waves that are being photographed, but one more thing is necessary for proof that it really was electrons that were diffracted. Electrons moving at high speed generate X-rays very easily; might it not have been these stray X-rays that caused the circles? This is easily tested, because electrons are strongly deflected by a magnet, whereas X-rays are unaffected. If then a magnet is brought near the apparatus, circles due to electrons will be shifted, but not so those due to X-rays. Thomson therefore took a photograph in a magnetic field, and proved that he really was diffracting electrons, because, as shown in Plate III, B, his system of circles was deflected into a new position.

The diffraction of electrons is a tremendous fact. To see its implications we had better translate the experiment into the simpler terms I have suggested above. Suppose that we have two very small holes in a shutter and that a beam of electrons all of the same speed falls on them. On the farther side we put a photographic plate. If it were practically possible to do the experiment on the appropriate scale, we can be quite certain that the plate would be crossed by bands of alternate light and darkness.

In order to summarise what we have found, we will imagine now that we take the old-fashioned view that an electron is a little speck of matter; but we will suppose that we have available the means of making holes in a shutter as small as we please, and photographic plates capable of recording the very feeblest effects. With this apparatus we start experimenting on the passage of electrons through holes. We begin with a single rather large hole, and observe that the stream of electrons goes right through in a straight line, and makes an image on the plate which is just the shape of the hole. If we look a little more carefully we see that there is some small effect also in the region of shadow, and we explain this, being conservative, by saying that some of the electrons graze the edges of the hole and so are scattered. Incidentally this was the earliest explanation offered of the diffraction of light. Next we make the hole smaller, and as we do so, we see that more and more electrons are scattered out of the direct line. This is also not unnatural, because in proportion as the hole gets smaller its edges will become more important. When the hole is very small indeed we are made a little unhappy about this explanation, because the central beam disappears altogether; perhaps we console ourselves with the explanation that the electrons are finite in size, and there is so little room for them to get through that practically all hit the edge and are scattered. But now we take two of these small holes close together, and at once find an effect that defeats our explanation completely. There are now places where no electrons come, whereas if one of the holes is blocked up, at once electrons begin to come there. The only possible way of explaining this is to say that each of the electrons knows all about both holes, or has gone through *both* holes at the same time, because only thus could we get the cancelling effect characteristic of interference. This is the direct inference from our experiment, and there is no escape from it. It is so contrary to all our ordinary ideas about matter, that it must be regarded as the greatest revolution that has happened in the whole of physical science.

CHAPTER IV

THE UNCERTAINTY PRINCIPLE

We have seen that direct experimental evidence pointed to electrons being waves, in the sense that when we send a stream of them through two holes we can only explain the result by supposing that, like a wave, each electron goes through both holes. We saw, moreover, that if a patch of wave-disturbance in a medium never encounters small obstacles it keeps together as it travels, and behaves in this way like an individual, which is what we think the characteristic of a particle. So we might at first sight be tempted to think that we had got a quite satisfactory and complete view of the character of an electron merely as being a wave of very short wave-length. But a little consideration shows that this will not do.

In the first place we have seen that though a patch of disturbance travels along as an individual with the definite group-velocity, there is always a region round its edges where the disturbance is slowly spreading. There is no way in which a wave can escape this gradual diffusion, and it means that ultimately it will become spread all over space. The rate of diffusion is smaller the larger the volume over which the waves are spread, so that it would be very slow for matter in bulk, and such waves would keep together a considerable time, but still they would not do so for ever. Even if we regarded the world as originally created in well-defined "wave-packets," they would certainly by now have spread indefinitely. We may say that the existence of fossils which have preserved their form unchanged for several hundred million years disproves the adequacy of the wave theory.

But the matter is worse than this, since we can do other experiments which seem immediately to disprove the validity of the wave theory. There exist substances which have the property of *scintillation* when struck by electrons. A scintillating screen is made by lightly powdering a sheet of glass with zinc sulphide crystals; when one of these crystals is struck by an electron it emits a faint spark, which can be seen in the dark with the help of a magnifying lens. When such a screen is exposed to a stream of electrons, scintillations appear irregularly all over it. The natural inference from this experiment is that the stream is like a shower of rain falling on the screen, and each scintillation is produced when a single drop hits the screen. We seem to have a perfect and complete proof that the electrons are little bullets each travelling along a line from source to target.

It looks as though we had arrived at a flat contradiction. This experiment tells us that the electron is a bullet moving along a line, while we could not explain Thomson's experiments without supposing that the electron went through two holes at the same time, as only a wave can do. To bring out the contradiction still more strongly we may combine both experiments into one; though this experiment has not actually been done, there is not a shadow of doubt what would be found if it were possible in practice. If we sent out a stream of electrons through two small holes close together, and then looked for scintillations, we should find these still appearing as isolated sparks, but the sparks would all occur in certain bands, and none at all in between at the places where the diffraction theory predicts darkness.

But if we afterwards block one of the two holes we shall destroy the interference and shall get scintillations everywhere. The crude way of saying what has happened is that the electron stream was a wave when it was going through the holes, but has miraculously turned itself into particles when it hits the screen. Of course such a description is not to be tolerated, since it would imply foresight on the part of the electron as to what was expected of it. We can imagine for instance that we could swindle the electron by pretending we were going to put a shutter with holes, so that it should get ready to be a wave, and then put a scintillating screen instead. Absurdities of this kind show that we have arrived at a very fundamental difficulty.

The elucidation of this contradiction is really the central point of the new quantum theory. The explanation, due to Heisenberg and Bohr, starts by showing that in fact the properties are not contradictory but complementary. Whatever the thing is that we call matter, it can be submitted to various experiments, some of which are devised to show wave properties, and some particle properties; but if we devise an experiment which shows the wave properties, that experiment debars us from observing the particle properties at the same time, and vice versa. Suppose, for instance, that we wanted to make sure that it really was particles that were going through our holes. We should set a scintillating screen over one of them, and whenever we saw a scintillation we should say that there was a particle coming to that hole. But in doing so we should have prevented the particle going through and so obviously should not get interference on the other side. The experiment, designed to show that the electron is a particle, has ruled out the possibility of our testing its wave character. Next we might try to improve the experiment, by making our screen so thin that the electron could produce a scintillation on it and still get through. Could we not then get interference between this part of the electron wave and the part that went through the other hole unimpeded? We should fail, because though the electron wave has got through the first hole, the mere act of exciting the scintillation will alter the phase of its wave, and if this phase changes there can be no interference. We have laid a trap for the electron to induce it to tell us which hole it went through, but when the electron answers the question by scintillating on its way through one hole, we can see that it cannot be expected to do the interference which would confess that it went through both.

In the understanding of such difficulties, as in the whole of the atomic theory, the name of Bohr will always stand pre-eminent. He was originally responsible for the theory of spectra, which proved to be the key to the quantum theory, and though many others made contributions of the first importance, it was mainly his critical judgment that guided the subject to the point where the wave character of matter was discovered. Not content with this he has continued in the same way ever since, and one of his greatest contributions has been the clear recognition that there is no conflict between the ideas of wave and particle. A situation arose rather like that in the early days of the discovery of relativity. The great idea which Einstein contributed to scientific philosophy was the principle that if a thing is essentially unobservable then it is not a real thing and our theories must not include it. He showed how the idea of absolute time was of this nature, and the whole beautiful structure of relativity was built up from that basis. But a selfconsistent mathematical formulation of the theory is not enough; it is also necessary to convince ourselves by examples that in fact it really is impossible to determine whether two events in different places occur at the same instant. We learn to understand the theory much better by "shamming stupid," trying to lay traps for the theory and seeing how it escapes from them. Much the same state of affairs has arisen in the quantum theory; we have considered one case where we laid a trap for the electron, trying to make it tell us whether it was wave or particle, and we have seen how it avoided the trap. We must convince ourselves that no experiment can be invented which should at the same time require the electron to behave like a particle and like a wave. The guiding principle which establishes this result is called the Uncertainty Principle, and we shall discuss this and with its help shall see how the conflict between wave and particle is always avoided.

As we have seen, some experiments with electrons exhibit their particle characters and some their wave characters. We cannot avoid thinking about both, and it is a very confusing thing to have to do. In one picture the electron is a little speck of dust, or a bullet, and in the other it is, shall we say, like a stormy sea, and it is not easy to see much resemblance between the two. There have been attempts to regard it as a speck of dust in a stormy sea, but it can definitely be said that they are no use at all. Perhaps the best description can be made by the use of a commercial expression; an electron is a particle "and/or" a wave. We must be ready all the time to think of it as either or both, but we must not mix the ideas. There are two half-worlds, each of which gives a partial view of the whole world; they are related to one another and interdependent, but they are expressed in different languages. We call the two half-worlds the "particle aspect" and the "wave aspect." There is nothing of the same kind anywhere else in scientific thought, but the absolute separation, yet interdependence, can perhaps be compared to a similar separation in metaphysics. There is a close interdependence between the objective thing that we see or hear and our subjective sensation of sight or hearing, and yet the two use wholly different languages. When a string on the piano vibrates 256 times a second, we hear "middle C" without any conception that there is anything happening 256 times; and when an æther vibration with twenty thousand waves in each centimetre strikes our eye, we see yellow, an ultimate sensation giving no hint of a wave motion. In the same way, when we burn a finger in the objective world it is because the atoms of the fire are moving about a little faster than those in ourselves, but actually all we feel is that it is too hot. There is the same kind of interrelation without identity between the wave and particle aspects of matter. It is tempting to carry the analogy a little further, and to decide which way round it is to be taken. I think that it is best to regard the wave aspect as analogous to the objective world, and the particle to the subjective; for example, we have a very direct and intimate perception of what a particle means if we are hit by a bullet, and on the other hand we have no intuitive knowledge whatever that light and sound have anything to do with waves. We may illustrate the analogy by making a table in two columns. The upper half describes the association of object and subject, the lower of wave and particle. I have anticipated the results of some later paragraphs in this lower part of the table.

Objective.	*Subjective.*
256 Air Vibrations a sec.	Middle C.
20,000 Æther Waves in a cm.	Yellow.
High Temperature.	Burn.

Wave Aspect.	*Particle Aspect.*
Wave-length.	Momentum.
Frequency.	Energy.
Number of Nodes.	Angular Momentum.

In making this comparison I must insist that the whole thing is only an analogy, and perhaps some will say a fanciful one. I am too bad a metaphysician to judge of this.

We must consider a little more closely the interdependence of the two aspects. In the last chapter we saw that it had been shown that under certain conditions the wave of an electron would have wave-length about the same as that of X-rays, that is about a hundred-millionth of a centimetre. Thomson and others have experimented with electrons which, regarded from the particle aspect, have various speeds, and have found that the wave-length is inversely proportional to the speed; but the limitations of experimental technique prevent the investigation of any very wide range of speeds. Theory, however, clearly indicates what the relation will be between speed and wave-length; indeed, the experimental work was really only a verification of the theory. The relationship was first given by de Broglie, and involves the *quantum* itself. The quantum is a certain universal constant which is always turning up in atomic theory. It was discovered and named by Planck in the opening year of the century, but we need not enter into his work here. That it is a perfectly genuine quantity is shown by the fact that it has with some precision the value $6{\cdot}545 \times 10^{-27}$ gr. sq. cm. per sec., but this does not really help anyone to understand it. Its nature is best described by saying that it is the single universal connecting link between the particle and wave aspects. The rule for finding the wave-length of any particle is to divide the quantum by the momentum of the particle, and this gives the ultimate meaning of the quantum. The rule is true not only for electrons, but also for protons, atoms, molecules, photons, and other bodies.

In order to observe the wave aspect easily, we want to get long waves, and that means small momentum, and small momentum can be got either by having low velocity or else very light particles. For this reason most experiments on the diffraction of particles have made use of electrons, the lightest particles that exist. It is interesting, however, to note that recently the diffraction of whole atoms has also been observed. We will consider a few of the associated values of speed and wave-length for electrons, but in doing so, it must be strongly emphasised that we are describing the two irreconcilable aspects of matter as though they could be mixed together. When I say that such and such a speed implies such and such a wave-length, it is only to be taken formally. It means that if we have a suitable grating, lateral spectra will be found corresponding in position to that wave-length.

In Thomson's experiments the electrons were set in motion by an electric field of about 20,000 volts. This gives them the high speed of $8{\cdot}5 \times 10^{9}$ cm. per sec., more than a quarter of the speed of light, which would carry them right round the earth in half a second. The associated wave-length is $0{\cdot}8 \times 10^{-9}$ cm., about a twentieth of the distance between the atoms in the analysing crystal. In Davisson's experiments much lower voltages were used. With 200 volts the speed would be a tenth as great and the wave-length ten times as much, nearly as great as the size of the atoms of the crystal. For much lower voltages the experiments would become very difficult, both because the electrons produce hardly any observable effect and also because they will not be diffracted when their wave-length is greater than the interatomic distance. It is very probable that these difficulties will be overcome in time; indeed, a beginning has already been made in that it has been found possible to observe the diffraction of electrons, though still rather fast ones, with an ordinary optical grating instead of a crystal. Though experiments are lacking, theory predicts with confidence the wave-lengths associated with slower electrons. An electron moving at the speed of a rifle bullet has wave-length about a thousandth of a millimetre, a length visible in a microscope. An electron moving at the rate of ordinary human walking would have wave-length of a size just visible to the naked eye; and one moving at the rate of that well-known standard of velocity, the snail's pace, would have wave-length several inches long.

These relations are so far only formal. We do not expect to be able to see the crests of the waves or anything of the kind, but are only maintaining that certain experiments, at present quite impracticable, would reveal diffraction effects which would imply these wave-lengths. But let us take the relationship more literally and see what it implies. An electron-particle moving at a rate of one centimetre a second is an electron-wave of length seven centimetres. Now a wave of seven centimetres does not by any means signify a wave with only two crests seven centimetres apart; it means a train of harmonic waves stretching to infinity in both directions, with all the crests regularly arranged at intervals of seven centimetres. Where then is the electron-particle? The answer is that it may be absolutely anywhere! This was the key to the elucidation of the whole quantum theory; it was entirely unforeseen, and it is the central fact of the new conceptions of matter.

Let us examine the question in more detail. We are trying to understand the connection between the two aspects of the electron by seeing what can be deduced about the electron-particle from our knowledge of the electron-wave. Perhaps we have taken a rather too pedantic view when we say that the mere calculation of a wave-length implies that there was an *infinite* train of harmonic waves, for after all a train of waves with twenty or thirty crests travels for a time in much the same way and could show diffraction. Such a group or *wave-packet*, as it is often called in the present connection, travels along with the group-velocity, but spreads a little as it goes. Fig. 17 is a sketch of a wave-packet and shows how the disturbance moves and spreads with the lapse of time. Where is the electron-particle with regard to this wave-packet? The medium carrying the electron-wave is undisturbed outside the two dotted cross-lines, and so we can say that the electron-particle is at all events somewhere between them, but we do not know whereabouts. The packet moves with the group-velocity, and the electron must keep in the packet, so it must move at something of the same rate too. But now there enters the important point that a wave-packet always spreads, and so at a later time is longer than at the start, and therefore the region available for the electron-particle is growing. This can be expressed in another way; we may say that the speed of the particle is not exactly the same as the group-velocity of the waves, but may be a little more or less. For example, if the particle moved with the front end of the packet, it would go faster than the group-velocity. If the motion of the packet is the only available guide from which we can deduce the speed of the electron-particle, then on account of the spreading of the wave-packet there will be an uncertainty in its speed. The point of the new outlook is that though we think of a particle as associated with the wave, it is *impossible* to know where in the wave it is, and *impossible* to say exactly how fast it is moving. Our first tendency is to resist this conclusion and to say that we can imagine ways of finding where the electron really is and how fast it is moving. We shall consider this point soon, and show how such an experiment is always defeated, but it will be best to accept it for the moment, simply taking the rule that the electron-particle is somewhere in the wave-packet, and consider what degree of uncertainty of position and motion this implies.

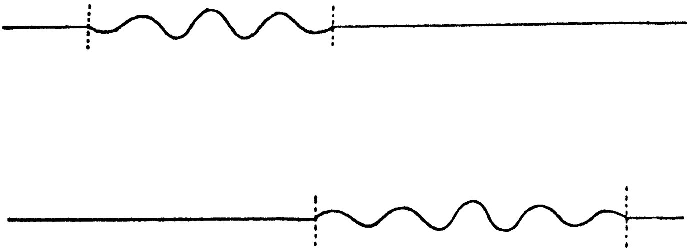


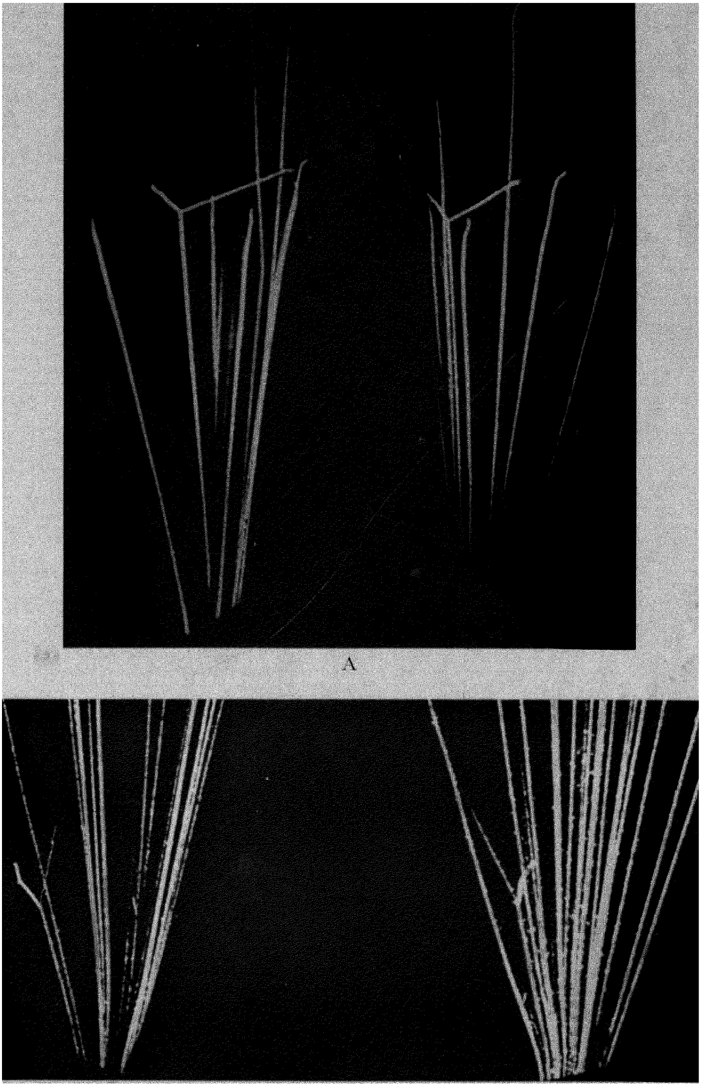
FIG. 17.—PROGRESS OF WAVE-PACKET

The uncertainty of position of the electron depends on the size of the wave-packet, so that for a long packet, containing a great many crests, the position of the particle is very uncertain. Such a wave-group on the other hand does not spread very rapidly, and so we can say that the velocity is rather precisely given. Next consider the opposite case of a very short wave-packet. This is illustrated by Fig. 8, which showed the propagation of a transverse wave along a bar; indeed, the actual propagation for a bar is the same as that for an electron. In the case of such a short packet the spreading is very rapid, so that the velocity is very uncertain. The general result thus is that the greater precision we demand for either position or velocity, the less the precision that can be assigned for the other. The rule is more definite than this, and can be given a rough numerical value. The product of the uncertainties of position and momentum of any particle cannot be brought below a value equal to the quantum. This is true for all particles, electrons, protons, photons, atoms, and so on. It is the Uncertainty Principle.

The relation between wave-length and momentum is only one way in which the wave and particle aspects are connected. There is another which in many ways is quite as important, and which must be described. We may recall that the character of a harmonic wave depends on both wave-length and wave-velocity, and that from these two a third can be devised, the frequency, which is the number of oscillations per second described by the medium at a fixed point. The frequency is really the most fundamental of the three, for if the medium has variable properties the wave-length and wave-velocity will vary in different places, but the frequency will be the same everywhere. The frequency of an electron belongs, of course, to its wave aspect, and the corresponding quantity in the particle aspect is the energy. The energy can be derived by multiplying the frequency by the quantum. There is also an Uncertainty Principle for the energy, just as there is for the momentum. This asserts that if we want to measure energy accurately we must take a long time in order to do so. If, on the other hand, we want to know the energy at a certain moment, we must obviously only use a short interval of time round that moment to do the measurement, and the value we obtain will be inaccurate. It is easy to see why this should be so by taking account of the wave aspect. An accurate knowledge of the energy implies an accurate knowledge of the frequency, and this knowledge can only be attained by letting the oscillation run through a great many cycles, that is to say, by taking a long time.

We have seen how the Uncertainty Principle arises quite naturally from the behaviour of wave-packets. But we must now assure ourselves that no experiment can be devised which would directly determine both position and speed with a higher accuracy than the principle permits; for if this could be done, then the particle's motion would have knowable characters not depending on the wave aspect, and this would mean that there must be a particle theory not involving the wave aspect. In that case *the whole of our present theory would collapse.* Now what are the most favourable conditions for testing the Uncertainty Principle? In the first place a simple calculation shows that bodies of ordinary size, on account of their great weight, have so little uncertainty of velocity, that the ordinary disturbances of the world will far exceed it. The effect only becomes perceptible for particles as light as atoms, and the most favourable case of all is the lightest particle, the electron. Let us therefore imagine that we have a sceptical experimenter, who refuses to believe in the wave theory, and sets to work to show that he can fix the position and speed of an electron at the same time with as high accuracy as he pleases. To make his experiment easier he will take the electron to be at rest, but it should be mentioned that this has nothing to do with the Uncertainty Principle; for that principle the difference between an electron at rest and moving at a centimetre a second is just the same as the difference between

B



ALPHA-RAY TRACKS

Two photographs of the tracks were taken from different directions, so that the tracks could be reconstructed in three dimensions. (*Photographs by Mr. P. Blackett*)

A. In the forked track the α-particle has met the nucleus of an oxygen atom and has been strongly deflected to the right. The oxygen nucleus has recoiled to the left.

B. One of the α-particles has met a nitrogen nucleus and has thrown out a proton from it, while itself remaining in the nucleus. The longer track in each photograph is the ejected proton, and the other the recoiling nucleus.

one moving at a thousand centimetres a second or a thousand and one. Our experimenter claims to have got an electron precisely fixed and at rest. We will cross-examine him about his work and see what he has found.

Q. How did you know the electron was there?

A. I saw it.

Q. An electron is a pretty small thing, and not easy to see. How did you manage?

A. I had a microscope.

Q. Even a microscope can only see things of the size of a wave-length of light. You can't be much of a precisian if you say you knew exactly where it was from that. I thought you said you would guarantee to know *exactly* where it was.

A. Yes, but you see I had taken a course in optics at the university, and so I was not caught out as easily as that. I invented a special X-ray microscope. It uses a wave-length of a thousand-millionth of an inch. With that I ought to be able to fix it close enough for you. Of course there are the cosmic rays with still shorter wave-length, but nobody seems to know where they come from, so they would not be very handy. Anyhow, I think I have done fairly well.

Q. Well, I haven't yet heard of an X-ray microscope on the market, but I suppose there will be one soon. Perhaps it would be pedantic to want you to do better. What did you see?

A. It was rather tiresome to get it going, but when I had done so an annoying thing happened. I knew the electron was there or thereabouts, because I had put it there; and it was at rest because otherwise it would have gone off while I was getting the microscope ready. Well, I was adjusting the microscope, and the electron was coming into focus beautifully, when it seemed to give a jump and run away. So that experiment was spoilt and I had to start again.

Q. Did you have better luck next time?

A. No. It was most curious; exactly the same thing happened every time. I think there must be something wrong with the microscope stage. I am going to have a shot at improving it. But as the microscope was certainly right in principle for seeing things to a thousand-millionth of an inch, and as the electron stayed there all the time I was focussing, it seems to me that I must be right. It is only a matter of overcoming the troublesome details that turn up in all experiments.

Q. It is not a matter of troublesome detail, and there is nothing wrong with your microscope stage. Your trouble is not with the electron being there and staying there, it is with the seeing of it. You can't see the electron without light to see it by, and the light disturbs the electron and drives it away. It does not matter how many different experiments you design, you will always get caught out in one way or another. There is no escape from the Uncertainty Principle.

The old particle theory breaks down not because it is inconceivable to imagine a particle at rest at a definite place, but because every method that can be contrived to observe that it is there always introduces a disturbing element. The ordinary experiments with gross matter are made with instruments so designed that they do not perceptibly disturb the object measured. It would be a poor way of measuring the length of a stick to hit it with a steam hammer, and if we want to see what a microbe looks like we do not place it in the focus of a powerful burning glass. The measuring instrument is always chosen lighter or weaker than the object measured; but this cannot be done when the object is the lightest thing that there is, an electron. In designing the experiment which is going to observe the electron, we have to examine all its details so as to be sure that the method of observation is not going itself to introduce some disturbance. We do not, of course, expect anything as crude as the burning of the microbe, but we must estimate what effect there may be. We shall find that the effect exactly explains poor A's troubles, but in order to do so we must make a digression.

It was known as early as the eighteenth century that all forms of wave exert a pressure on any obstacle that is reflecting them. This can easily be seen with a stretched string, as in Fig. 18. Instead of tying the string to a support at the right-hand end, suppose that it passes through a hole in a frame and is made fast somewhere beyond. The string just fits the hole, and the frame is held firm. When a wave of vibration travels along from the left towards the frame it cannot pass the hole but is reflected back, forming "standing waves" with the hole as one of the nodes.[1] Now consider the forces acting on the frame. The string has a bend at the hole (except at the moments when the phase of the wave makes it straight), and the frame has to bear the pressure of this bending. The direction of the force evidently bisects the angle between the string on the two sides of the hole, and so is nearly sideways, but not quite so, for the bisector must always fall to the right a little way behind the plane of the frame. The principal component of the force which is in the line of the frame is alternately in opposite directions, and so averages out; but the longitudinal component is to the right whether the vibrating part of the string is up or down, and so on averaging there is a residual force to the right. If we do not wish the frame to move, it will be necessary to hold it with a small force pushing towards the left, the direction from which the waves are coming. This means that the waves exert a pressure on the frame. More detailed considerations show this pressure to be proportional not to the amplitude, but to the intensity of the waves.

---

[1] See p. 37.

It is found to be a universal rule that waves of every kind exert a pressure on an obstacle reflecting them. This must therefore be true of light, and the effect was predicted and many of its consequences were worked out long before the phenomenon was observed. The effect is very small indeed for any available source of light—the total force exerted by the sun shining vertically on a square mile of the earth is equal to a weight of about 3 lb. The first attempt to detect the effect had a rather surprising result. Crookes made a little windmill with vanes blackened on one side and polished on the other. The polished sides reflect light, while the black absorb it. In consequence the force on the black side is half as great, for though it is receiving the wave it is not returning it. When exposed to a bright light,

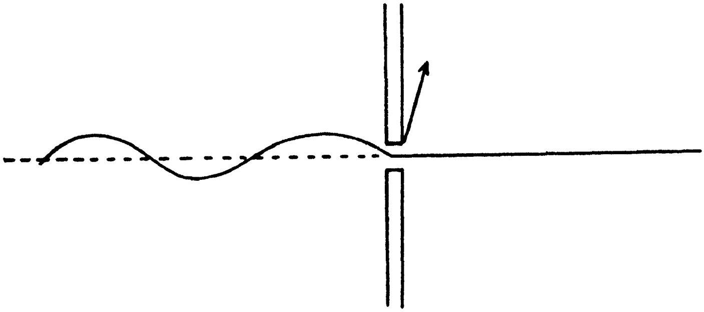


FIG. 18.—PRESSURE OF WAVES

The string passes through the hole in the frame. It exerts a force on the frame in the direction of the arrow. As the string vibrates this is sometimes up and sometimes down, but it always falls behind the frame and so tends to push it back to the right.

the radiometer should therefore go round with the black sides leading. It does go round, but the other way! This was ultimately traced to an effect of the irregular heating of the residual gas in the vessel; though very small it still far outweighs the minute direct effect of the light. It is only in comparatively recent times that this difficulty was overcome, first by Lebedev in Russia, and the actual pressure of light directly measured.

The pressure observed in this way is the gross pressure acting on the whole bulk of a body. This must be regarded as the result of all the separate pressures on the atoms and electrons. The simplest inference that could be made was that each electron just took its proportional share of the whole. But with the development of the quantum theory it became possible to admit that this might not be so. If for example a few of the atoms got a violent kick and the rest none at all, the cohesive forces of the material would enable the few to drag the many with them, and the result in bulk would be just the same as though all the atoms had experienced a feeble force. This was the guiding idea in the very important discovery made by A. H. Compton in 1922. From general considerations of the quantum theory as it then was, Compton put forward the idea that when light falls on an electron the process should be regarded as though it were a collision between two particles. Remember that this was before anyone dreamt of the wave aspect for matter, and though the particle aspect of light was well known, no one before had ever dared to take it in anything like as literal a form. With the details of the Compton effect we shall be concerned in a later chapter, here it suffices to describe the outline. When an electron scatters light it is thereby caused to recoil, and the speed of the recoil depends only on the wave-length of the light and not at all on its brilliance. For visible light the recoil is feeble, but for X-rays it becomes very easily perceptible, and in fact Compton verified his theory in all its details by using X-rays. The only distinction between the effect of a bright light and of a faint one is that the bright light will scatter an electron sooner than the faint, but the speed at which the electron goes will be the same in either case, provided the wave-length of the light is the same.

We may now return to our experiment with the microscope, and we know where the trouble lies. A microscope system consists of two parts, the condenser and the microscope proper. The condenser focusses light on the object, the object scatters it, and the microscope then refocusses the scattered light into the eye. If we are to see an object, that object must have scattered light, and must itself recoil in consequence. So the mere fact that we see the electron guarantees that it is set in motion; even if it was at rest before we saw it, it cannot be so afterwards. The mere carrying out of the experiment spoils the result aimed at. Notice that if we are content with knowing the position rather inaccurately, we need not use light of a very short wave-length, and shall not then get much recoil; but if we want the position accurately, we must have a short wave-length and then the recoil will be large. So we see that the Uncertainty Principle is maintained; high precision in position or velocity can only be attained by the sacrifice of precision in the other.

We have seen how one method of defrauding the Uncertainty Principle is overcome, but may there not be others that are more successful? Of course the only way of proving that none can succeed is by the use of the general principles of the quantum theory, but it is profitable to consider a few further examples and show in detail how the attempts fail. We have seen that a microscope is no help, and so we try to make use of a method that does not require one. If, for example, we have a shutter with a very small hole in it, and have a source of electrons on one side, then if we find an electron on the other side, we know it must have come through the hole, and so we can locate its position in that way. We must work out the experimental arrangements a little more carefully. The experiment might be done in the way shown in Fig. 19. We have a pair of parallel plates *ABC* and *FGH*. Electrons start at rest from *ABC* and fall under the influence of a force towards *FGH*. They will move exactly in directions parallel to *AF*. The idea behind the experiment is a little more complicated than before, because it is necessary to consider the different directions separately. When an electron has emerged from the hole at *G*, it thereby tells us what its position is as far as concerns the direction *GF*, but says nothing about its position in the direction *GM*. So our interest is in the velocity in the direction *GF*, and we do not care what the component in the direction *GM* may be. Now to the left of *G* we know the electron's motion to be along *BG*,

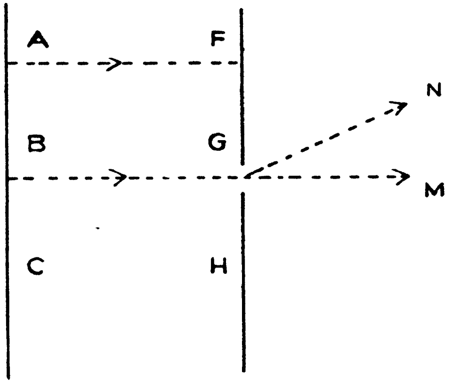


FIG. 19.—FIXING POSITION OF ELECTRON BY SLIT

that is to say, its component in the direction *GF* is zero. We seem to have conquered the Uncertainty Principle since we know the speed to be zero, and the position is given as accurately as we please by making the hole at *G* small enough. But we only know that the electron was on the line *BG* and not on the line *AF*, because it emerges at *G*, and in emerging it will be diffracted, say along the line *GN*, and so will acquire a component of velocity transversely, and one which is uncertain in amount. Once again the mere fact of finding the position has introduced an unwanted velocity. Notice, too, that if the hole is rather large, there is not much diffraction and so very little uncertainty in the velocity, but to counterbalance this advantage there is no very precise knowledge of the position; while on the other hand, with a very small hole we can fix the position accurately, but pay for it by strong diffraction and so great uncertainty in the speed of the electron after it has emerged.

We will not yet confess defeat. It is true that the electron has been diffracted, but can we not measure through what angle it has turned? If we can do so, we can conquer the Uncertainty Principle, not by avoiding the effect of the observing instrument, but just as successfully by measuring it. We might proceed for example as follows. The electron has altered its course in passing through *G*. A force of some kind must be necessary to produce this deflection, and this will react on the shutter and tend to push it in the opposite direction. If then we measure this reaction, we can assert what path the electron has taken, and this is what we want to know. The simplest way of observing the reaction is to make the shutter free and very light, so that as the electron passes it will be set in motion. We adopt this method. But if the shutter is free, how do we know where the slit is at the moment the electron is passing? We have settled the question of the momentum satisfactorily, but in doing so have lost the position. We must try again, and devise a plan by which to know the position. We therefore send a beam of light through the hole, and by watching this beam we can see where the hole is. Surely we now know both position and momentum at the same time. But no, we have forgotten something, for the light itself will behave in the same way as it did in the microscope; it will be diffracted at the hole and will itself start giving impulses to the shutter. There is no way of knowing whether the impulse we observe belongs to the electron or to the light, so that we have regained the measurement of the position but have paid the price by once more losing the momentum.

It is not by any means easy always to detect the fallacy in experiments like this, but there always is something wrong. Each time we find the defect in our process we must install some extra piece of apparatus to put it right, and the addition, in the course of overcoming the old difficulty, always introduces a new one. There is no escape from the Uncertainty Principle.

Here it may be well to summarise the argument of the chapter. Consider the old-fashioned assumption that an electron is a pure particle. This implies that it is "really" placed at a precise spot, and "really" has a precise speed. The assumption works for most ordinary occurrences, but totally fails to explain the diffraction of electrons. We therefore inquire what evidence there is for the assumption. If it is to mean anything at all to say that the particle is placed at a precise point, it must mean that we think we could devise methods of *accurately* finding where that point is. Now from considering various experiments we have seen that there is a definite limit to the accuracy with which position and velocity can be simultaneously determined. Hence there is no warrant in experiment for our assumption that the electron is a pure particle. We take instead the conception of the two aspects, wave and particle. The particle aspect is what we observe, for instance by a scintillation, but the wave aspect underlies the occurrence and determines in what regions it is possible that the particle should be observed. We took the case of a wave-packet, and saw that the position and velocity of the particle-electron would be fixed exactly to the degree admitted by experiment, if we made the hypothesis that the particle is equally likely to be observed anywhere in the region of the packet. This hypothesis removes all the contradictions between wave and particle, and makes a consistent theory. There is one respect in which the statement has hitherto been incomplete, since the relation of wave and particle has only been described for a wave-packet. It is necessary to extend the theory to the more general case where the electron-wave is not in an isolated packet, and it is easy to do so. The *probability* of observing the electron-particle at any point is proportional to the *intensity* of the electron-wave there. This is the general rule governing the relation of wave and particle.

The Uncertainty Principle is essentially only concerned with the future; we can install instruments which will tell us as much about the past as we like. Suppose, for example, that we have two shutters each provided with a very small hole, and a source of electrons to the left of both. The holes are usually blocked up, but for a very short space of time I first open the one in the left shutter, and at a definite time later I do the same for the one on the right. I look for electrons to the right of both shutters. If I see one I can be quite certain that it went along the line between the holes and took a definite time in doing so—that is to say, I can know its position and speed precisely. What the principle asserts is that this knowledge is no use in predicting what is going to happen later, for it gives no knowledge of how the electron will be diffracted on emerging from the second hole.

This must revolutionise our ideas about one of the most fundamental principles which has always been accepted in science, the principle of causality. We are accustomed to take it for granted that a *full* knowledge of the present would enable us confidently to predict the future. When we are defeated in our attempts at prophecy we attribute it to ignorance, with the tacit assumption that with more knowledge of the present we could have done better. It never occurred to anyone that the present is definitely unknowable; but we have just seen that the mere effort to know it cannot help introducing new errors in the determination. We used to be faintly ashamed of the fact that we were not omniscient, but now ignorance has become respectable.

It has been suggested that the new outlook will remove the well-known philosophical conflict between the doctrines of free will and determinism, and it has been welcomed by many for that reason. I would personally offer a most strenuous opposition to any such idea. The question is a philosophic one outside the region of thought of physics, and I cannot see that physical theory provides any new loophole. If we are to find room for free will within the realm governed by physical science, we have to suppose that the motions of our own bodies are in some way free not to obey the inexorable commands of the older mechanics. At first sight it might appear that the Uncertainty Principle provides the necessary latitude, but this is contradicted by closer consideration. We cannot say exactly what will happen to a single electron, but we can confidently estimate the probabilities. If an experiment is carried out with a thousand electrons, what was a probability for one becomes nearly a certainty. Physical theory confidently predicts that the millions of millions of electrons in our bodies will behave even more regularly, and that to find a case of noticeable departure from the average we should have to wait for a time quite fantastically longer than the estimated age of the universe. How then does the Uncertainty Principle help to free us from the bonds of determinism?

CHAPTER V

THE ATOM

In the last chapter we were discussing the closest possible approximation that the wave theory would allow towards regarding the motion of an electron as though it was an ordinary particle of the kind with which we are familiar. We discussed either pure harmonic waves, or groups which were approximately harmonic, but we must now consider the question of the analysis of waves in general without this restriction. In view of its greater familiarity we will return to the case of light, though nearly everything would apply equally well to any other kinds of wave.

We described the action of a diffraction grating for harmonic waves by showing that the first lateral spectrum was formed in the direction where the circular wavelet from one slit falls just a wave-length behind that from its neighbour.[1] But there is another way of regarding the whole matter which is most instructive. We need no longer consider harmonic waves, but may take a wave of any form whatever. The simplest possible form consists of a single pulse, as in Fig. 20. On the incident side of the grating there is then one crest advancing, and the medium is undisturbed except just near it. When the crest reaches the grating each slit sends out a wavelet consisting also of a single crest. Now take any position *P* on the farther side and observe what will happen. The first disturbance will occur when the wave from *A* arrives; at a definite later time that from *B* will arrive, at an equal later time that from *C*, and so on. The observer will receive a wave consisting of a set of regularly spaced crests equal in number to the number of slits of the grating. But this is very like a periodic wave; indeed, it is approximately periodic, and the observer will see the light coloured. If the observer had taken up a position at *Q*, at a wider angle, the successive pulses would have been farther apart and the wave-length would be longer. Thus when such a pulse of light passes the grating, there appear on the other side waves of different wave-lengths travelling in various directions. The incident pulse may be regarded as white light, and on the other side blue light (which has shortest wave-length) will appear close in to the central beam, and farther out the other colours of the spectrum down to the red. It is perhaps right to

---

[1] See Fig. 12, p. 64.

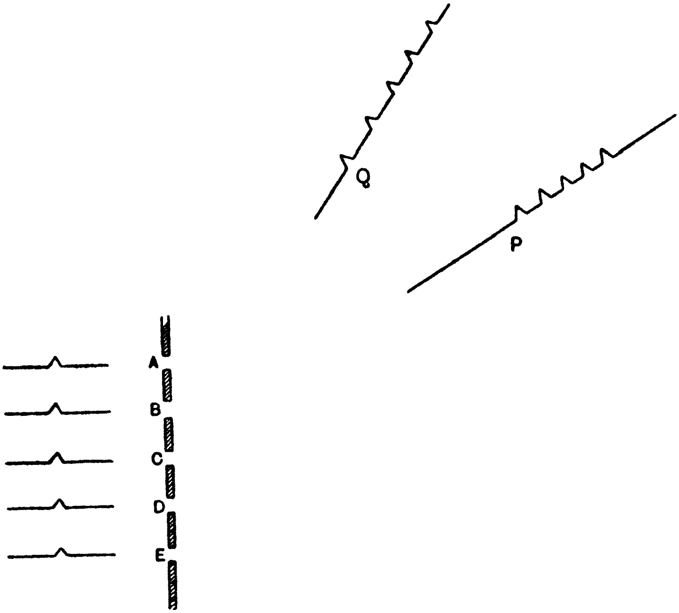


FIG. 20.—PULSE PASSING THROUGH GRATING

An observer at *P* will receive a succession of pulses at equal intervals. The grating has converted the non-periodic pulse into a periodic process. An observer at *Q* will receive the succession of pulses at wider intervals than at *P*, so that if the light at *P* is blue, that at *Q* will be red.

add, though the matter need not trouble us, that, though the waves are periodic, they are not harmonic. This means that each colour is not pure, but has the colour of its octave, etc., superposed on it, but as the whole visible spectrum is less than an octave, the effect will not be noticed by the eye.

The interesting point about this way of regarding the action of a grating is that we started with a pulse having no periodic characters, but have ended with periodic waves. It would evidently be right to say that the grating has manufactured the colours, and not as we said before, that it selects already existing colours and sends them in a particular direction. A good many years ago there was a controversy on the subject of the nature of white light. Some said that white light was nothing but a mixture of all the colours, while the opposite school maintained the view I have just suggested, that it is the grating which manufactures the colours. The controversy was settled in the best possible way by the final agreement that both parties were right; a single pulse can be regarded as a mixture of harmonic waves of all lengths, as may be shown by mathematical analysis. The argument reduced itself to a question of taste. Nevertheless it contains an important lesson, for it serves to remind us that the harmonic analysis is a matter of convenience, not of necessity. For light this is clear and easy to accept, but it is much more curious for an electron-wave. An electron-wave which is not harmonic will show diffraction in a grating just like white light.[1] If there is really only a single electron, we see where it goes, and, being rather prejudiced by our habits of thinking of the electron as a particle, we are apt to think that even before it came to the grating the electron had a definite speed and so was "really" a harmonic wave with wave-length corresponding to that speed. But this is going far beyond the evidence, and it is just as right to say that it is the grating that has manufactured this wave-length and therefore this speed. For consider how we could test the question. We should have to measure the speed somehow before it came to the grating, and this mere determination of its speed implies that it will have been made into a harmonic wave by some means or other. In fact, we cannot test what its speed is without spoiling the purpose of the subsequent experiment. Whatever we do we are begging the question.

---

[1] There is a slight difference, because the electron-wave changes form as it goes, so that the simple geometrical construction of the pulses is not available; but this is unimportant.

The point I want to make is that the analysis of waves into their harmonic components is really a matter of convenience. The harmonic wave is the type which possesses the simplest mathematical expression in certain very important circumstances, in particular for free space; and it is the type that many of our experimental instruments naturally give. But under other circumstances a different analysis may prove more suitable, and we are now going to consider one of these.

We have so far only discussed the behaviour of a free electron in open space, or at most have limited its motion by such negative things as shutters and slits, which either stop it or let it go on undisturbed. But this is, of course, a very incomplete view of things, just as we get a very imperfect idea of light when we overlook the fact that it may be refracted, as for example when its direction is changed in passing from air to water. In the same way we must study what happens to matter in the presence of a *field* of force. Here the work of Hamilton on Least Action steps in and provides a rather definite answer.

We recall that Hamilton worked out very exact analogies between the behaviour of rays of light and particles of matter. He showed that one could construct the parabola of a projectile by the same process as one calculates the path of rays of light through a medium, using in both cases the Principle of Least Action. It is true that the medium in question would not be one favoured by the optician, because the speed at which the light goes must vary from point to point in a manner that would defeat the most skilful glass-maker. Still, the fact remains that if a glass could be constructed dense in some parts and rare in others, according to a suitable rule, it would send the light in a parabola, exactly imitating the path of a projectile under gravity.

Hamilton did not carry his methods further than this, because he can hardly have dreamt that the reason why his analogy was so perfect was that the things described were exactly the same. But de Broglie pushed the matter to its logical conclusion by saying that if light and matter are *refracted* in the same way, then they ought to be *diffracted* in the same way too. In other words, we have got to carry the analogy to the end, and whereas Hamilton worked out the case of rays of light and rays of matter —a circumlocution for particles—so now we must work out the analogy for waves of light and waves of matter.

Now the things which will exert forces on our electron are other electrons and atoms, and these must naturally have the same wave properties as the one of which we are studying the behaviour. We shall have to consider this full problem later, but can make a start by taking a mixed description in which we regard the other atoms, etc., as though they were fixed *particles* exerting given forces on the special electron which we treat as a wave. This procedure, like the description of nearly everything in physics, is inexact, but it is drawn rather in the way that an artist draws a portrait, reproducing the fine details of the face of his victim, but treating the background with much less accuracy.

The most important motion to study is the motion of the electron in an atom. The very simplest case of all is the atom of hydrogen, which contains a single proton for nucleus and our electron. We represent the attractive force which the fixed nucleus exerts, by supposing that the electron is like a ray of light moving in a refractive medium. Hamilton's work gives definite information as to how the refraction varies with the distance from the centre. It is very slight at great distances, and grows to an enormous value near the nucleus. If we could make a glass having these properties, we should find that rays of light are very strongly bent round towards the centre, in fact so strongly that they may be trapped, so that they go round and round it for ever. The refraction is in fact such that the path of the light would be an ellipse, described in exactly the same way as the planets describe their paths round the sun. There is nothing wonderful in this; it happens because we have arranged the refraction so that it should happen, and this we have done because we know that the force of electric attraction is of just the same type as that of gravitation. If the electron were really a little speck of electricity it would certainly describe this path.

But now we have to bring in the point which has been concerning us so much; that the idea of the electron as a simple particle is a crude approximation which only works when the obstacles are large, and must be replaced by the wave aspect when the obstacles become small. An atom is a very small thing, and so we must expect that we have to study it in proper detail. This means that we must proceed to make the analogy not with the crude theory of rays of light, but with the accurate theory of waves, involving a study of the progressive spreading of the rays as they travel. This was one of the main ideas in the work of de Broglie; it was clothed in a strict mathematical form by Schrödinger, and it was Schrödinger's work on the problem I am now discussing that first really brought success to de Broglie's general but rather vague conceptions.

We now suppose that we have a group of waves somewhere near the nucleus. They constitute a wave-packet and travel at first very much as would the rays of the particle model; that is to say, the centre of the packet goes round in an ellipse. But now we have to remember that a wave-packet always spreads as it goes, with the result that after a number of revolutions the packet will have extended more or less right round the centre. How are we to describe conveniently what it will be like? A new process of analysis for the description must be found, and there exists one which is very convenient, and which gives us all that we want.

Consider a new analogy. When we ring a bell, the clapper displaces a small piece of the surface outwards. This piece drags the neighbouring parts after it by virtue of the rigidity of the metal, and so the motion is handed on. In the meanwhile the first piece, after moving outwards a certain small distance, stops and returns past its original position and continues to vibrate to and fro. So does every other piece of the bell, and a motion of very great complication ensues. If we want to describe this motion, the most primitive way of doing so would be to say where each particle of the bell was at every instant of time. But this would be terribly complicated and would contain a great deal of information that does not interest the musician, who merely wants to know what the bell sounds like. He is concerned only with the question of what note the bell has, and whether it is harsh or sweet. The actual study of bells shows that they always give out a number of tones (which as a matter of fact are usually quite discordant); the pitch of the bell is a sort of average of the louder among these tones, and its *timbre* depends on the relative strengths of the various tones. The most useful analysis has nothing to say as to how the bell moves when struck by the clapper, but yields the pitch and intensity of these tones. The analysis is made into what are called the "normal modes" of vibration. When the bell vibrates in one of its normal modes, it passes through a succession of definite shapes of deformation, and every piece of it executes a harmonic vibration with all of them in the same (or opposite) phase. The main question in the study of a bell is to know the form of these modes, and the pitch of the associated pure tones. When these have been found we may say that the problem has been mastered; it is true that there remains outstanding the question of how loud the tones will be relatively to one another, when the clapper strikes the bell in some particular place, but once the modes are fully known this is comparatively easy to find.

We will now examine a little more closely the nature of the analysis into normal modes, which is the essential part of the discussion of every vibrating system. We will first take a simpler case than a bell, that of a violin-string. This is much the simplest that there is, and indeed is too simple in some ways to be of use for the general discussion. We have already considered how a train of harmonic waves travels along an unlimited string, and have also seen how they would be reflected at a fixed end into *standing* waves. But we now have two fixed ends and we want a different analysis. We want to know the tones given by the violin, and for this we study what particular shapes the string assumes when it gives out a single tone. These are shown in Fig. 21. The first gives the fundamental tone; the second is the octave with half the wave-length; the third the "twelfth" in musical parlance, and the fourth the double octave and so on. The wave-lengths are respectively, half, third, quarter and so on of the fundamental, and the sweetness of the violin's note is due to the exact relationships of pitch which exist between these tones. They are usually described by their numbers 1, 2, 3 . . . simply in order of their decreasing wave-length, but there is a more geometrical description which is of greater utility for our present purpose. A *node* is a point on the string which does not move during the vibration,[1] and we see

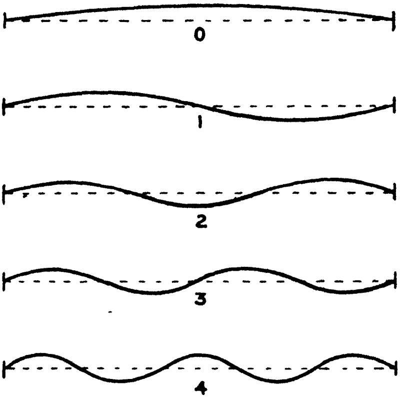


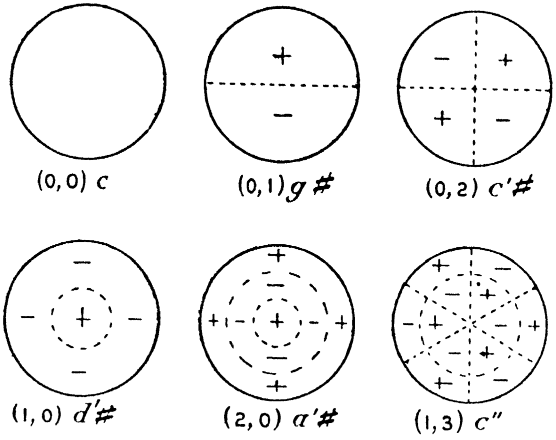
FIG. 21.—NORMAL MODES OF STRING

The attached numbers give the number of nodes. The pitch of the second mode is the octave of that of the fundamental, that of the third is the "twelfth," of the fourth the double octave, and so on.

---

[1] See also Fig. 3, p. 37, which is essentially the same question as the present.

that the first mode has none, the second one, the third two, and so on. Though in fact in the case of the string it is not usual to count in this way, it is in general a very good description of the modes of vibration to number them off by the number of nodes they have—here as 0, 1, 2 . . . . We shall apply the method to some other cases.



(0,0) *c* (0,1) *g*♯ (0,2) *c*′♯

(1,0) *d*′♯ (2,0) *a*′♯ (1,3) *c*″

FIG. 22.—NORMAL MODES OF DRUM

The numbers indicate the character of the vibration. The first of the pair describes the number of nodal circles, the second of nodal radii. The motion is such that when the parts marked + are up, those marked – are down, and vice versa. The drum is supposed tuned to *c*, and the pitches of the other modes are indicated.

We will pass to two dimensions, and consider the vibration of a stretched circular membrane, which is illustrated in Fig. 22. The nodes are now no longer points which do not move, but lines. As the drum vibrates in one of its modes, a nodal line on it stays still, and when the part on one side of the line is up, that on the other is down. We must consider some of the modes of vibration. The fundamental mode possesses no nodes, the whole membrane rises and falls together, and the motion is largest at the centre. Next there can be a symmetrical mode with a single circular nodal line, so that when the middle is up the parts nearest the edge are down and vice versa. Unlike the violin-string this mode is not in tune with the fundamental; the interval between them is about a minor tenth. Then there is another mode with two circular nodes and so on. Next there are unsymmetrical modes of motion. In the first there is one straight nodal line across a diameter so that when half is up the other half is down. The pitch of this is at an interval of about a minor sixth above the fundamental. Then there is one with two nodes at right angles, then one with three, and so on. Next we may have both radial and transverse nodes of which examples are shown. We can devise a quick and easy way of describing the various modes, by taking two numbers, the first of which stands for the number of circular nodal lines and the second the number of straight radial ones. Thus the fundamental is 0, 0 and the last one 1, 3.

When we come to a three-dimensional system, of which a sphere containing vibrating air is an example, we shall get nodal surfaces instead of nodal lines. These may be either spheres, or else planes or perhaps cones through the centre. A nodal surface is a fixed surface which is not crossed by the air-particles during the motion. We require now three numbers to describe each mode. The first gives the number of nodal spheres, the second of cones cutting the surface in circles of latitude, and the third the number of planes cutting the surface in circles of longitude. It is not easy to make a diagram which clearly shows a motion in three dimensions, but if we limit ourselves to the case where the third number is zero, the motion is symmetrical round an axis, and so can be seen by showing the motion on a section through the centre. This is done in Fig. 23. The dotted lines are the nodes, and the arrows roughly indicate the maximum motions of the air-particles. It will be seen that the motion is rather complicated, but corresponds roughly to a compression of the air into certain of the corners between the nodal surfaces and a rarefaction in the others. Half a period later the compressions and rarefactions will be interchanged.

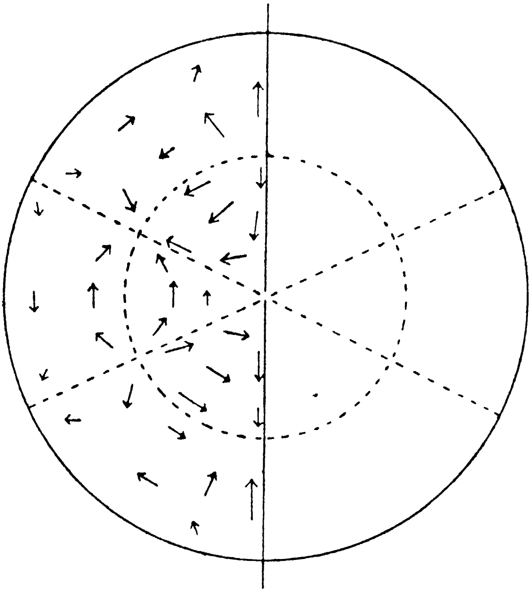


FIG. 23.—VIBRATING SPHERE OF AIR

In this mode the motion is symmetrical about the vertical axis, and so may be described by means of the motion of the air-particles in a section through that axis. The mode shown is (1, 3, 0). There is one spherical node, three nodes in cones of latitude, and no nodes in planes of longitude. The arrows roughly show the motion of the air-particles.

This digression tells us how to attack the problem with which we started, the passage of our electron-wave round the nucleus of an atom. We started with a wave-packet near the nucleus, and this we liken to the blow given to the bell by the clapper. The packet is strongly bent by the refraction and spreads as it goes, just as the disturbance by the clapper spreads as it travels round the bell. Finally in both cases the motion becomes very complicated if we think about what is happening at a particular place in the atom or bell, and a new analysis is found more appropriate. We analyse into the normal modes, each of which is independent of the others. We

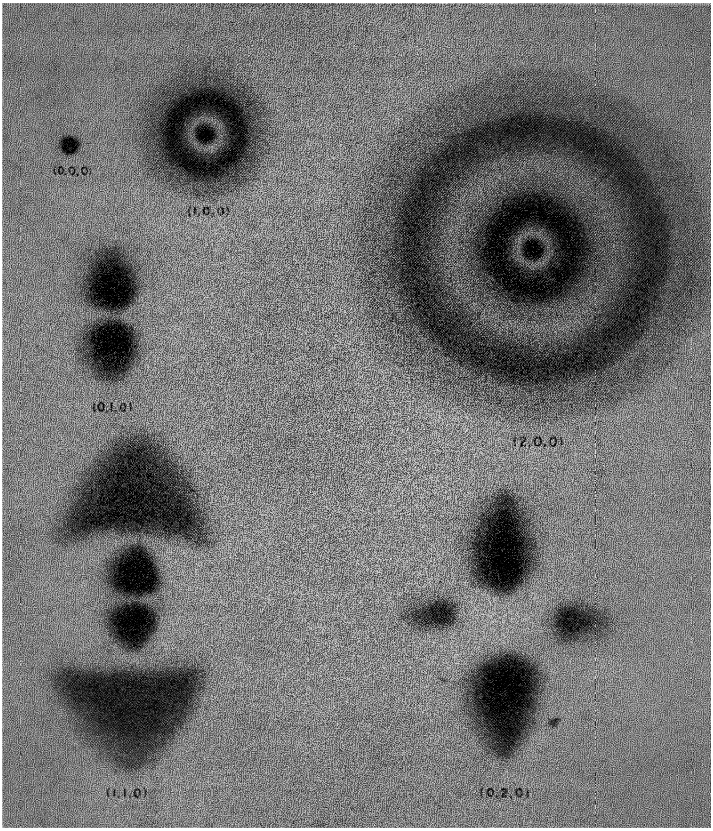


FIG. 24.—MODES OF HYDROGEN ATOM

The diagrams show the *intensity* of vibration at each place, and so indicate the probability of finding the particle-electron there. Each is supposed to be rotated about a vertical axis through its centre. Thus (0, 2, 0) would consist of a ring round the equator and two large lumps round the poles.

describe these normal modes by attaching to each three numbers to represent the nodes in the motion. The main distinction between this case and the sphere of air is that for an atom there is no definite outer boundary. That does not make any real difference, because the refraction is so strong at the centre that the system is still closed, in the sense that the electron never escapes, even though it is closed *by infinity* and not by a finite bounding wall.

It is not easy to make diagrams exhibiting three-dimensional properties, but the main characteristics of some of the modes can be represented by the device of cutting a section across the centre and blackening the figure in proportion to the intensity of the wave. A few of these are sketched in Fig. 24, but it may be mentioned that the nodal numbers attached to them are not quite the same as those which are in prevalent use in describing the modes.[1]

In discussing vibrations in general we have seen that every normal mode is associated with a frequency of vibration which is special to it; and so a very essential, indeed the chief, part of the description of the vibrations consists in finding their frequencies. In the present case these obey a rather simple law, which is worth giving. The rule is this. Take the first two numbers of the three which describe the mode, add them together, and add on 1.[2] Square this number; the result is proportional to the period of the vibration, so that its reciprocal is proportional to the frequency. The fundamental mode is 0, 0, 0. This gives $(0+0+1)^2=1$ by our rule. The mode (1, 1, 0) gives $(1+1+1)^2=9$, and so its frequency is oneninth of the fundamental. It must be emphasised here that these are not the frequencies of the emitted light, which follow a rule to be described later; but the frequencies of the vibration can be deduced with great accuracy from those of the emitted light. They consist in numbers of fifteen figures, implying a thousand million million vibrations a second, and it is characteristic of the precision of spectroscopy that the first eight of the fifteen are known exactly. The frequency of the fundamental mode is usually described in a slightly indirect way, by imagining that it gives out a wave of light corresponding to that frequency (which it does not in fact do), and counting the number of wave-lengths in a centimetre. This yields a much more manageable number of six instead of fifteen figures. It is called Rydberg's wave number,[1] and is 109737·4, implying that there would be about a hundred thousand wave-crests in each centimetre. Rydberg's wave number can be calculated from a knowledge of the numerical values of the charge and mass of the electron, the speed of light and the quantum, but it can be measured with far higher precision than any of these quantities, and so serves as an important check on their values.

---

[1] The usual numbers are denoted n, l, m. If our three numbers are a, b, c, then $l=b$, $n=a+b+1$. As regards m the modes are chosen in a slightly different way from that described by c; m may take all integer values between $\pm b$.

[2] The result is the number n, described in the last footnote.

We have been discussing the natural frequencies and modes of the electron in the atom, making use of the analogy of a bell. The analogy is exact up to a point, but it must not be pushed too far or it will give wholly wrong results. It is correct as long as we regard the bell, and the atom, as closed systems without external connections. But the purpose of a bell is to make a noise, and our interest in the atom is just because it can exert effects outside, so that an essential part of the problem is to consider the linkage between either vibrating system and the external world. For the bell that is easy. As the surface of it moves in and out it pushes the air in front of it to and fro, and thus it sends out a wave of the same frequency as the vibration of the bell, so that when we know the normal modes of the bell we can at once say that it will emit sound of the corresponding frequencies. But for the atom this is not so; the linkage is of a much more subtle kind. There is only an indirect connection between the frequencies of the vibrating system of the atom and the frequency of the light emitted by it. The full discussion of this question belongs to a subject we shall take up in a later chapter, the interaction of particles (here an atom and a photon), but without such deep consideration as is there given, a fairly good account can be given of it in another way.

---

[1] This assumes the nucleus of the atom to be fixed. For actual hydrogen there is a perceptible change because the proton is not infinitely heavy compared to the electron.

We have so far dealt with the wave aspect of the electron in the atom, but its particle aspect is also important, and provides a picture which is often useful. It was discovered by Bohr twelve years before the wave aspect, and Bohr's Theory of Spectra was historically the origin of all the developments of the new quantum theory. At the time of Bohr's discovery the electron was nothing but a little electrified piece of grit. How does it behave in an atom? The law of electric attraction to the nucleus is just the same as the law of gravitational attraction, and so the electron will describe the same track as that of a planet round the sun. We therefore anticipate that the electron will describe an ellipse with the nucleus in a focus. But there is a fundamental difficulty about this, for a planet can describe an orbit of any size, whereas an atom is known to be of a quite definite size. The principles of dynamics, as exemplified in astronomy, are therefore inadequate to explain the atom, and a new principle must be imported from outside. This new principle was the fundamental achievement of Bohr's theory; it is called quantisation. In order to understand it we must examine the orbits more closely.

The orbit of a planet is usually an ellipse, but may take the simpler form of a circle. In his earliest work Bohr assumed the orbits to be always circular; later on the restriction was removed, but we need not go into that. When a planet moves in a circle its motion may be specified in a variety of ways. The most obvious is to give the radius of the circle, for since the force of attraction is known, the speed and time of revolution can all be determined from this. But there are other alternative ways of specifying the orbit, by making use of dynamical instead of geometrical quantities. In planetary motion there are two quantities which are conserved, as we saw in the first chapter; one is the energy and the other is the angular momentum. If the value of either is given, all the characters of the orbit are fixed by it. Bohr adopted the principle that the angular momentum must be a multiple of the quantum.[1] The consequence is that the electron may only go round in one of a set of definite circles. If the angular momentum is a single quantum, the radius is called "the radius of the first hydrogen orbit," and is about $0{\cdot}5 \times 10^{-8}$ cm., or about a two-hundred-millionth of a centimetre. This length has a better claim than any other to be the ultimate footrule of the world. The other permissible orbits have radii 4, 9, 16 . . . times as great, corresponding respectively to angular momentum 2, 3 4 . . . times the quantum.

This principle of quantisation was the essential point of the theory, and it had no rational basis, but could be justified by its enormous success. We can now understand why it worked. In free space the thing which in the wave aspect was the wave-length of the electron became its momentum in the particle aspect. There is a similar correspondence for the motion in the atom, and it turns out that what in the wave aspect is the number of nodes becomes the angular momentum in the particle aspect. Fig. 25 shows one particular normal mode (it is the mode 0, 5, 5), represented by drawing a section through the axis, and it will be seen that if the figure is rotated about the axis it represents a circular ring. This is a hazy picture of the circular orbit which the electron-particle would describe, and it can be verified that if the number of nodes is counted and multiplied by the quantum, there results an angular momentum equal to that of the particle in its orbit.

---

[1] Strictly speaking, of the quantum divided by 2π; that is to say, it is to be a multiple of $1{\cdot}04 \times 10^{-27}$ gr. sq. cm. per sec.

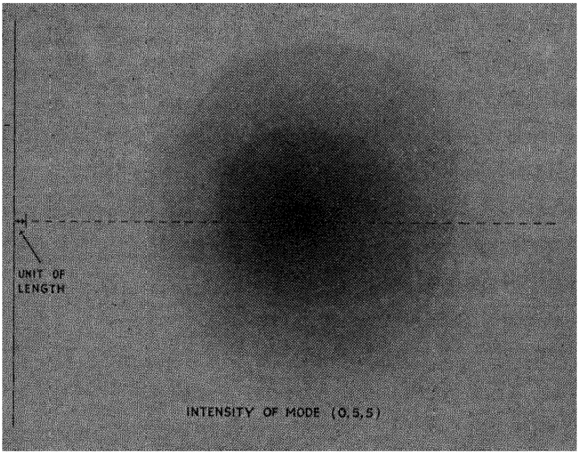


FIG. 25.—A MODE OF THE HYDROGEN ATOM

The diagram is to be rotated about the axis on the left. The result is a hazy ring, the wave aspect of one of Bohr's orbits. The unit of length shown is the " radius of the first hydrogen orbit."

The wave aspect improves on the hard simplicity of the particle aspect by smudging out its detail.

We can understand the association of the two aspects in another way, which is practically that suggested by de Broglie. Consider a wave-packet moving on the circumference of a circle fairly far from the nucleus. If it were quite free it would be regarded as being the wave aspect of an electron moving with a certain speed; we will suppose that this speed is just the one which would in the presence of the nucleus make the particle describe the circular orbit. The wave-packet will then travel round in this circle, but will spread slowly as it goes. It will ultimately spread right round the circle in a hazy ring. After the head has overlapped the tail their effects must be superposed, and this will depend on their phase relationship; unless they fit together in phase they will cancel out with one another. So we analyse the packet as the superposition of the types of wave which will not cancel out, and this means that we regard it as a set of waves in one of which there are 5 crests in the circle, in another 4, and in another 6, and so on. The process of quantisation becomes automatic.

There is an important difference between the wave and the particle description suggested by consideration of this wave-packet. As long as the electron was a little piece of grit it was axiomatic that it could only describe a single orbit. For the wave that is not so, since a vibrating system can oscillate simultaneously in a number of its modes. The wave aspect of an electron is a vibrating system, and so we must admit that the electron may be simultaneously in several normal modes. If we like slightly to misapply a term borrowed from the particle aspect, we may say that in the new quantum theory, as opposed to the old, an electron can describe several orbits at the same time. The advantage is all on the side of this view, even though many experiments seem to show that an electron must be in a single orbit. The contradiction can be reconciled in the same way as was described at the beginning of this chapter in connection with a free electron going through a grating. There we saw that it was perfectly justifiable to say that it was the diffraction of the grating which gave the electron its definite speed. In just the same way here, when we ask the electron which mode it is in, it will give a definite answer; but it is the experiment observing the mode which itself tells the electron what it has got to do. We shall discuss an example of this in detail in the next chapter.

We must return to the old particle theory. The quantisation fixed the orbits completely, and therefore among other things their energy. By the use of a hydraulic metaphor each quantised state is called an *energy level*, and a sketch of the states of the atom can be made by drawing horizontal lines to represent the levels. This is done in Fig. 26. They are all negative quantities, and fall in the relative positions -1, $-\frac{1}{4}$, $-\frac{1}{9}$, and so on. These levels were originally worked out by Bohr using the old quantum theory, but they are also given by the wave theory; there, of course, they occur as frequencies, since frequency is the wave aspect of energy. Now comes the second main point in Bohr's theory. The atom would always, like water, try to go to its lowest level. In doing so it would have to lose energy, and this it would do by radiation. If then it started in the second state it must lose the difference of energy between $-\frac{1}{4}$ and -1, that is $\frac{3}{4}$, and this will travel away as a photon of energy $\frac{3}{4}$. In the wave aspect this signifies that the light will have wave-number $\frac{3}{4}$ that of Rydberg's wave-number. This radiation from hydrogen was found long ago by Lyman, and all the other principal changes of level give radiations which have also been found. This way of describing what light the atom emits is that which is nearly always used, and is much the most convenient practically. It serves to bring out the difference between the atom and the bell, but on the other hand it does not seem to accord at all well with the account of the atom given by the pure wave aspect. In a later chapter we shall take up this aspect, and show how it leads to the same results.

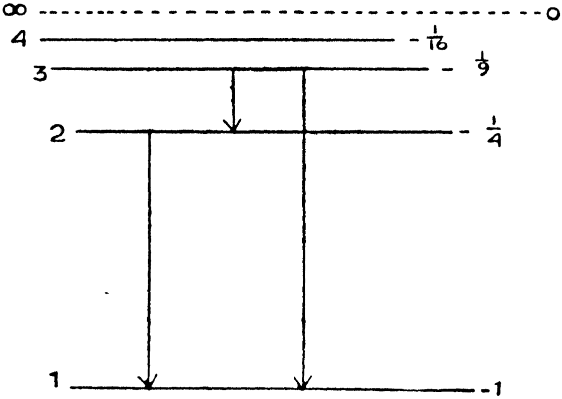


FIG. 26.—ENERGY LEVELS OF HYDROGEN

The numbers on the left indicate the angular momentum, those on the right the energy. The lowest level is that corresponding to the "first hydrogen orbit."

In conclusion it should be said that in my account of the hydrogen atom I have been making a contrast between the wave and particle aspects which does not do justice to the particle aspect. The description with particles was that of the old quantum theory, which had many defects. It is impossible to reconcile this crude theory with the wave theory without considerable modification, but it is possible to make that modification, and still retain the conception of particles in a useful manner. The accomplishment of this was due to Heisenberg, who invented a new dynamics, which could do with particles all that the waves can do, and was indeed soon recognised to be the same in substance as the wave theory, though very different in form. Heisenberg's dynamics is peculiar in that it breaks the ordinary laws of algebra, in particular the commutative law of multiplication, and for this reason it is rather difficult to follow. Many of the leading experts in the new quantum theory seem to prefer its methods, but for the purpose of gaining physical insight into what is happening, I cannot help thinking that a greater insistence on the wave aspect, helped out by a much cruder use of the particle aspect, is the best course.

CHAPTER VI

POLARISATION

IN the present chapter we shall be dealing with a rather different subject from anything discussed hitherto; we shall be concerned with what may be called the internal qualities of the three primitive things, light, electrons, and protons. We will begin with light.

One of the great optical discoveries of the seventeenth century was "double refraction"—the fact that objects seen through certain crystals look double. Into the intricacies of this effect we need not enter. The phenomenon was reduced to simple terms by Huyghens, and its character was summarised by Newton, who said that a ray of light could have "sides." To explain this curious phrase it will be best not to attempt a description of the experiments since all experiments on double refraction involve rather troublesome geometry. It will suffice to describe the theory derived from those experiments, which was given some century and a half later by Fresnel in the second great era of optical discovery.

In our study of light-waves in general we took a number of analogies, water-waves, sound-waves, a stretched wire, and so on, and all the properties deduced were equally true for all of them. But we must now specialise a little more, and in order to do so we may fix our attention on the vibrations of a stretched wire. Imagine that we have a wire stretched horizontally like a telephone wire. It may vibrate either up and down, or sideways, or else it may vibrate simultaneously both up and aside at the same time. The complete motion of the wire may be regarded as a superposition of the two types of vibration, each of which behaves as though independent of the other. We could, for instance, easily make a device to cut out the horizontal vibration and leave the vertical free to travel on alone; it is only necessary to imagine that the wire passes at some place through a slot which prevents its moving sideways there, but does not impede its motion up and down. To specify the vibration of the string we must therefore describe two things instead of one, both the horizontal and the vertical motion. Light has the same character; if we have a plane wave of light advancing say horizontally, it requires a double specification to describe it. Moreover, the analogy can be worked out more closely, for one of the two specified characters can be regarded as a vertical vibration and the other as a horizontal; and a well-known instrument, the Nicol prism, exists which, like the slot for the wire, can cut out one of the vibrations and leave the other to pass on unimpeded.

In double refraction the crystal on account of its lack of symmetry refracts the horizontal and vertical vibrations differently, so that they become separated as in Fig. 27, and the result is that whereas one ray entered the crystal, two will emerge from it. But the two represent vibrations, one of which corresponds to the wire moving vertically and the other horizontally. We may liken the incident beam to a round stick, and the two emergent beams to thin laths with their thicknesses at right angles to one another; that was what Newton meant when he said the light had acquired "sides." To remove possible misunderstanding, it should be made clear that the "sides" do not refer to the shape of the cross-sections of the emergent beams, which are just the same as that of the incident beam; it is an intrinsic property of the light, and is technically called polarisation. The observation of polarised light is usually made with the help of a Nicol prism, and we can understand it best by reverting to our vibrating wire. Suppose that this wire is carrying a vibration which we think is polarised and we want to know in which direction it is vibrating. We should pass the wire somewhere through a slot, and this would only allow vibrations to pass which were in the direction of the slot. We should turn the slot round until no vibration at all passes, for we then know that the incident vibration must have been exactly perpendicular. In the same way, if polarised light is to be studied, we hold a Nicol prism in front of the eye and twist it round. At one setting the light will be extinguished, and it must therefore have been polarised at right angles to this direction.

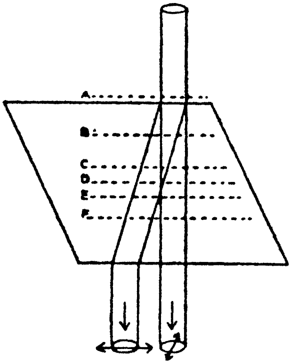


FIG. 27
DOUBLE REFRACTION
The incident beam is split into two by the crystal, and these exhibit polarisations mutually at right angles, as is indicated by the arrows at the bottom of the diagram.

There are more complicated types of polarisation which must also be considered, and the stretched wire will serve perfectly to illustrate them. Imagine that the wire is carrying simultaneously two harmonic waves of the same wave-length, one horizontal and the other vertical. A great deal then depends on how the two are related in phase. Suppose first that the phases are the same. Then when a point of the wire is at its maximum in height it will be at its farthest to the right, and at the moment at which it is passing through its mean position vertically it will be doing so horizontally also. It will thus vibrate, keeping all the time in some definite inclined plane; and therefore such a motion is called plane polarised. But there are more interesting cases. Suppose that the vertical and horizontal motions are equal in amplitude, but differ in phase by a quarter of a wave-length. When a point of the string is at the top, it is horizontally at its mean position, and when at the extreme right it is at its mean level. Tracing out its course, it is found to describe a circle, but there are two ways of doing so, either like the hands of a clock or the opposite way about. We thus have two types of motion called respectively right- and left-handed circularly polarised light. More general motions are also possible, in which a point of the string describes an ellipse, and these are therefore called elliptically polarised light. These types of polarised light are shown in Fig. 28. There remains to be described one other kind of light, and that is ordinary unpolarised light. Curiously enough this is more difficult to describe accurately than polarised light. A harmonic wave goes on for ever, and so if it is once polarised, say elliptically in any way, it will stay so. It follows that ordinary light cannot be in the form of an accurate harmonic wave. It consists of a nearly harmonic wave which changes its polarisation every million cycles or so. It takes on in turn polarisation in every different direction, and so, on the average, loses the "sides" which are the character of polarised light.

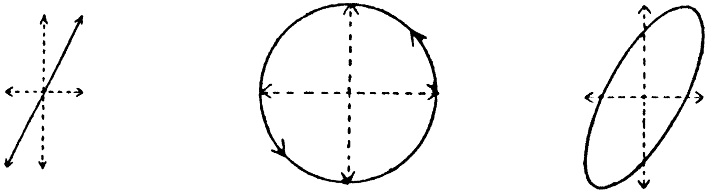


FIG. 28.—TYPES OF POLARISED LIGHT

A, plane polarised. B, circularly polarised. C, elliptically polarised.

The dotted lines indicate the amplitudes of the vertical and horizontal components.

The description which I have given of polarisation is only directly applicable to plane waves, that is to say, waves in which the phase is the same all over a set of parallel infinite planes, at right angles to which the waves are advancing. This is the only type of waves for which the idea of plane or circular polarisation has any very strict meaning, though it is possible to extend it to cases where the waves are nearly plane. The mathematical theory for three dimensions is completely understood, and gives no special trouble, but there are rather severe difficulties in making a physical model for three dimensions. The most natural one is to imagine the æther as a sort of jelly in which the waves are represented by supposing the particles of the jelly to be deflected sideways. This gives a fair rough model of what happens, but the model goes definitely wrong if taken too literally, so that it must not be insisted on. In thinking about polarisation it is usually best to pick out the direction in which the wave is travelling and to think of a wire stretching in that direction, and not to consider the complications of other directions.

We can now see what happens when a beam of unpolarised light enters a crystal. The wave-velocities of the two components are different inside the crystal, and the groups of waves consequently travel in different directions. We can represent what happens by cutting cross-sections of the beams at successive depths in the crystal. Sections have been cut at *AB* . . . *F* in Fig. 27, and the results are shown in Fig. 29. The circular beam is at first unpolarised, but the two components gradually separate. A little way inside the crystal the cross-section will be two overlapping circles. The part they share is still unpolarised, but the outside parts are plane polarised at right angles. These parts grow at the expense of the shared part, until finally there are two separate circles polarised at right angles to one another. These then travel on separately until they emerge from the other side of the crystal.

We must also look at the particle aspect of light, which, unlike that of the electron, is not so familiar to us as the wave aspect. A photon is unlike an electron in that it has no permanent mass, but it has a perfectly definite energy, which is proportional to the frequency of the light-wave in the way I described when discussing the Uncertainty Principle. It is this energy that heats a body when it absorbs the light. In the same way it has momentum, which is related to the wave-length of the light; it is this momentum that gives rise to the pressure of radiation. Energy and momentum are possessed by the particle aspect of any wave, but when we take polarisation into account the photon may also have angular momentum. Reverting to the stretched wire, imagine that we whirl one end of it round in a small circle. A circularly polarised wave will travel along and tend to give a circular motion to a body at the other end; this means that angular momentum is travelling along the string. Just the same is true for light; if the photon is circularly polarised it possesses angular momentum, and the amount is exactly a quantum. If the light is plane polarised it has none, and if it is elliptically polarised it has as its angular momentum some fraction

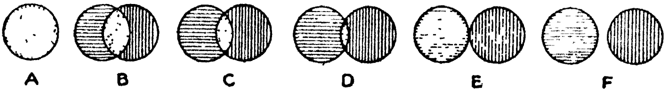


FIG. 29.—DOUBLE REFRACTION

The diagrams represent the cross-sections of the beams shown in Fig. 27. The dotted parts are unpolarised, and the vertical and horizontal shadings represent the two types of polarised light.

of a quantum, which is larger the more nearly the ellipse resembles a circle. We may summarise the properties of the particle aspect of light by imagining that a beam of circularly polarised light falls on a small black screen and is absorbed. The screen will then show three effects. It will be heated by the energy it absorbs; it will be pushed away by the momentum it absorbs; and it will acquire a twist on account of the angular momentum it absorbs. The first is a familiar phenomenon which we experience every time we sit near a fire, the second has been verified with the greatest difficulty, but I have never heard of any experiment yet made which exhibited the third.

We now turn our attention to the electron to see whether there is any effect there similar to polarisation. I shall present the matter in a different order from that of the actual discovery. We described the structure of the hydrogen atom from the wave aspect, by the idea that the electron-wave was vibrating in the manner of one of the normal modes. Fig. 25 showed one of these modes as a hazy ring surrounding the nucleus, and that mode will do for my present purpose. Let us now for the moment imagine that the circulating wave was not an electron-wave but a light-wave, and consider what the effect would be of allowing for the polarisation of the light. We can roughly picture what happens by imagining a wire lying in a circle round the equator, and stretched by being repelled from the centre. Such a wire could carry waves, but they would now be of two different kinds. In one the vibration would be in and out in the plane of the equator, in the other it would be at right angles, so that the motion of each particle of the wire would be north and south. The two types of wave are shown in Fig. 30. The same is true for the light vibrations; in one the light is polarised in the equatorial plane, in the other at right angles to it. But closer investigation shows that we have to give up any accurate analogy with the wire, for if the normal modes are worked out for light vibrations, it appears that the hazy rings which represent them do not quite fall in the same place; one has a slightly larger radius than the other. Furthermore, associated with this slight difference of position there is a slight difference of frequency. Thus we can say that if the electron-wave exhibited polarisation like light, it would have the number of its normal modes doubled, and they would fall into pairs close together in position, and close together in frequency.

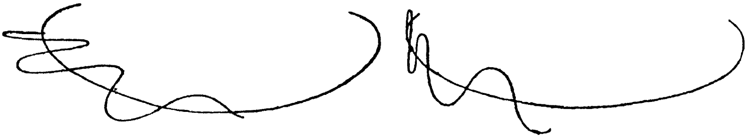


FIG. 30.—VIBRATIONS OF CIRCULAR WIRE

The two normal modes are shown in perspective. In the left-hand diagram the motion is in and out, in the right-hand one it is up and down.

This separation would exhibit itself to us through the spectrum of the atom, which would show twice as many energy levels as the number we should expect if the waves were not polarised.

Now this doubling of the number of energy levels in spectra was well known long before the discovery of the wave aspect of electrons. There is a curious sort of accident in the numerical values which conceals it in hydrogen, but the same effect is easily evident in other spectra. The atoms of the alkali elements, sodium, potassium, etc., resemble hydrogen in that they have a single electron comparatively free, which is responsible for all the optical and chemical effects. Unlike hydrogen they, of course, possess other electrons outside the nucleus —sodium has ten of them—but these are tightly bound and can be regarded as inert for most purposes. Their only effect is to alter the forces acting on the single electron, so that the particle motion would no longer be a simple ellipse; the normal modes will correspondingly be distorted and their frequencies altered. The spectrum of sodium is perhaps the best known of all spectra; it consists of a very brilliant yellow light, which is easily shown to consist of two wave-lengths close together. The rest of the spectrum is mainly ultra-violet, but it consists entirely of doublets of the same kind, and the same is true of all the alkali elements. There is definite proof that the occurrence of the wave-lengths in pairs is not an accident, but that the doublets are closely related together. This evidence is provided by the Zeeman effect, which is the name given to the phenomena which occur when an atom is caused to emit light while under the influence of a magnetic field. The Zeeman effect is very complicated in its details, and we need not consider them here, even though in fact it was this effect more than any other which led to the clearing up of the difficulties of the old quantum theory. It will suffice to say that according to simple theory the magnetic field ought to split a *single* frequency into a triplet of frequencies, but that in fact the effect of a very strong field on doublets is to break up the *pair* into a single triplet.

For a long time the doubling of levels remained a mystery, and it was explained from the particle aspect only a few weeks before the whole theory was transformed by the discovery of the wave aspect. We may consider the particle aspect first, for though we have seen that the idea of polarisation would explain the doubling qualitatively, it turns out that the polarisation is of a very different type from that of light, and it will be made clearer by the particle aspect that this must be so. Two young Dutch physicists, Uhlenbeck and Goudsmit, introduced the hypothesis of the "spinning electron." When an electric charge of any kind moves, it gives rise to magnetic forces. The most familiar case is that of an electric current in a wire, which gives the force which deflects the permanent magnet, which tells us by means of a pointer that the battery in the car is being charged. So, too, when the electron describes a circle round the nucleus, as we considered in the last chapter, it exerts a magnetic force and in fact turns the atom into a little magnet.

Now suppose that an electron is a solid sphere of electricity spinning about an axis. The parts round the equator constitute moving electricity, and so produce a magnetic force. As a result the electron behaves as though it was not merely an electric charge, but also as though attached to this charge there was a little bar magnet. In its magnetic character, though not its electric, it may be likened to the earth; for the earth rotates and also exerts towards the north pole the magnetic force which directs our compasses. We must consider how fast the spin is, and how strong the magnet, and for these magnitudes a study of the details of the Zeeman effect provides a quite definite answer. The spin is such that the angular momentum of the electron is just half a quantum, and the magnet has the strength of just one of the units called "Bohr magnetons." The magneton is the fundamental unit of magnetism; it is the magnet which is produced by an electron (supposed for the moment without spin) describing the "first hydrogen orbit."

With the help of the spinning electron and the simple particle theory we can now see how the number of energy levels is doubled. Consider one of the circular orbits.

There are two different ways that this may be performed, according to whether the electron spins in the same direction as that of its revolution round the centre, or in the opposite direction. The first case may be likened to the motion of the earth, which spins on its axis in the same direction as that in which it revolves round the sun. The second is rare in the solar system, the nearest example we can find to it is in the planet Neptune, the satellite of which circles round its primary in the opposite direction from that described by Neptune round the sun. The distinction in the rotations carries with it a difference in the magnetic effect. When the two motions are the same way about, the magnetic effect of the spin reinforces that of the revolution of the electron in its orbit, but when they are in opposite directions they partially counteract one another.

We must next consider the wave aspect of the spinning electron. As we saw, the idea of polarisation would account for the doubling, but it turns out that the type must be radically different from that of light. The easiest way of seeing this is to return to the two types of polarised light-wave that circulated in our model. We saw that in one of them the vibration was in and out, and in the other up and down. But to represent the spinning electron we want a different kind of wave, because we want to be able to discriminate between up and down. We need some sort of wave which can represent up and not down, or down and not up, whereas the light-wave cannot help combining up and down together. A pair of waves possessing the characters of up and down respectively can be found, but there is no easy geometrical way of describing them, and indeed the possibility of the existence of waves of this character was hardly suspected until the necessity had been felt for making a wave theory of the spinning electron. The correct formulation of such waves was finally due to Dirac, who derived their form by starting from the general principle that an electron could not exist without having magnetism in it, if it were to fit properly into the theory of relativity.

The electron in free space must also have polarised components, and here there is the apparent complication that no particular directions are up and down, so that the magnet can point in any direction. This can be overcome in the wave aspect by the principle of superposition, for a suitable combination of the up-wave and the down-wave can yield a wave corresponding to a magnet pointing in any direction. Let us imagine that we have a beam of polarised electrons, and compare its character with that of the corresponding beam of polarised photons. Both carry energy and momentum, as do all waves; but on account of their polarisation both may also carry angular momentum. For the light it is something less than a quantum, turning about an axis in the direction in which the light is travelling. On the other hand, the electron always carries exactly half a quantum, but it is pointing in any direction, and need bear no relation to the direction of motion.

The evidence which has been cited, that the electron-wave is polarised, has been rather indirect, depending as it does on considerations of the energy levels in spectra. Up to a point it has been verified much more directly, and with the usual disorderly chronology of discovery, it was in fact verified before it was discovered. The work that did this was a very celebrated experiment made by Stern and Gerlach in 1921. To describe it we must first consider how one can study the character of a magnet. If we have a piece of iron, and we want to know if it is a magnet, the easiest way is to hang it up freely, and see if it sets itself in some direction like a compass; the earth's magnetic field tends to twist any magnet to point along it. What happens is that the north pole of the magnet is pulled northwards by the earth's field, and at the same time the south pole is pushed to an equal extent southwards. As the two poles are bound together, all that the magnet can do is to twist round. Unfortunately there is no way of hanging up an electron by a thread, or of seeing how it points if we could hang it, so that a more complicated process is necessary. We must contrive to get a pull instead of a mere twist. This can be accomplished, if we can arrange that the north pole is pulled more strongly than the south pole is pushed, and to do this we must have a non-uniform field. Then the force on the north pole will attract the magnet a little more strongly than the force on the south pole repels it, and on the balance there will be a weak force which will tend to move the magnet. This is shown in Fig. 31.

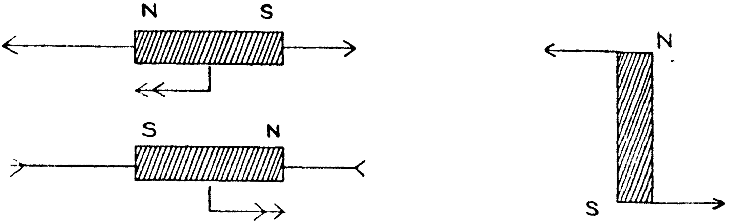


FIG. 31.—MAGNET IN NON-UNIFORM FIELD

The field is towards the left and stronger on the left. The single arrows show the force on each pole, the double arrow their resultant. In the right-hand diagram, though the magnet will tend to turn, there is no resultant pull towards the left.

The experiment of Stern and Gerlach consists in sending a beam of atoms through a non-uniform magnetic field. If an atom is a magnet pointing one way, the north pole will be more strongly attracted than the south is repelled, and it will move into the stronger parts of the non-uniform field. On the other hand, if it is pointing the other way, the south pole will be more strongly repelled than the north is attracted, and the net effect will be repulsion. If a magnet were pointing at right angles to the field, it might be expected that it would first twist round to point along the field and then be attracted. But this overlooks an important character of both the atom and the spinning electron, and that is that they are rotating like tops. When the force of gravity pulls a top downwards, the top does not fall down, but instead moves round horizontally in a motion of *precession.* In the same way, under the influence of the magnetic force the atom describes a precessing motion round the line of the force, and there is no tendency for its magnetic axis to shift towards that line. If then an atom enters the field crosswise, it should stay crosswise; its north pole would be attracted with exactly the same force as its south pole is repelled, and, though it will precess, it will go straight through the field undeflected.

The experiment was first done with silver atoms, though more recently it has been extended to a good many other elements. Silver is rather like sodium, in that it has one active electron, and the rest of its electrons, 46 in number, form an inert core round the nucleus. These atoms are boiled off from a small furnace in a very high vacuum, and therefore travel at a high speed in perfectly straight lines from the furnace. Slits select from them a single fine beam, which passes close to an iron knife-edge, magnetised so as to produce a non-uniform field. A plate of glass receives them at the other end, and the atoms slowly form a deposit on it. It takes many hours to get a deposit that is visible, but chemical means have been devised for saving time by developing much smaller amounts of deposit long before they have become visible. A sketch of the apparatus is shown in Fig. 32.

Let us now consider what we should expect to see. The atoms are little magnets, and will be pointing arbitrarily in all directions as they leave the furnace. Some will happen to be pointing straight in the direction of the magnet and will be attracted by it, others will be pointing away and will be repelled. We should also expect that most of the atoms would be intermediate, pointing obliquely, and precessing round at a constant inclination to the direction of the magnetic field. These would be attracted or repelled to a greater or less degree depending on their inclinations. Furthermore, we must consider the shape of the image transversely. The atoms passing the middle of the slit come opposite the knife-edge of the magnet, and here the field is very non-uniform and the force therefore is great, but at the ends the field is much more uniform, so that there is hardly any pull or push on the atoms that pass there. If we first do an experiment

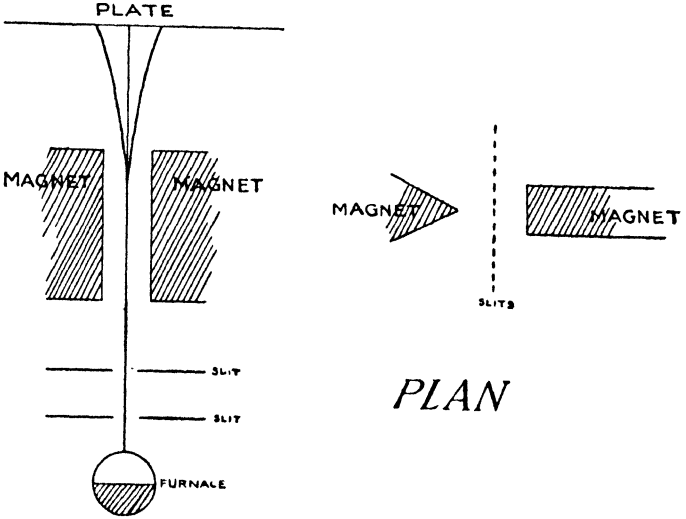


FIG. 32.—STERN-GERLACH EXPERIMENT

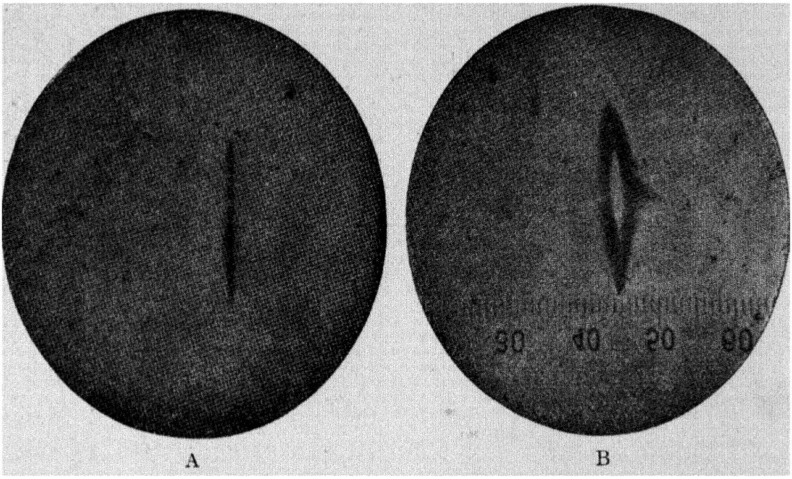
Silver atoms are boiled off from the furnace. They pass through the slits and between the poles of the magnet (shown also in plan), and fall on the plate.

without the magnetic field, we expect to get a long narrow line, which is just the line of the slit. When the field is put on, we expect that the ends will be much as they were, but that the middle will be broadened out. The result that this crude theory suggests is that we ought to see a lens-shaped smudge. When the experiment is carried out we get a lens-shaped image, but, instead of being solid, it is hollow! This result tells us that the crude picture we made was wrong; and that the atoms must be pointing half straight towards the magnet and half straight away from it, so that they all get the full pull or push of the magnetic field. Plate IV shows the result of one of the experiments of Stern and Gerlach.

If this experiment had been made twenty years ago it would have been regarded as almost incredible, but at the time it was made the actual result was anticipated, though perhaps with that uneasy feeling which we all know when our hope has a little outrun our faith. When the experiments were first made, the theory with which they agreed was comparatively crude. The idea of the energy levels of the atom and of electron orbits was known, and it was also known that in a magnetic field there would be two energy levels, one of which would be due to a magnet pointing down the field and the other up. But there was an element that could not be understood at all; for the atoms must be pointing in all directions before they entered the field—for how could they know what was coming to them—and then they must miraculously jump half one way and half the other.

With the advent of the wave theory of matter, the whole business became clear, and can be explained from the point of view which we used in discussing the diffraction of electrons. According to this principle we have to think about waves and also about particles, but

PLATE IV



THE EXPERIMENT OF STERN AND GERLACH

Silver atoms pass through fine slits, then past the poles of a magnet, and finally fall on a screen, condensing there and so producing an image A is the image produced when the magnet is not acting, and this simply shows that the atoms normally move in straight lines. In B the magnet is acting, and on account of the non-uniformity of its field the atoms are deflected The shape of the image shows that the atoms are magnets, half of them pointing up and half down the field In the printed scale twenty divisions correspond to a millimetre, so that the actual image is very small (*By courtesy of* '*Zeitschr f Physik*,' ix, p 350)

the ideas are complementary to one another and must not be mixed. We can simplify the statement by regarding the core of the silver atom as merely ballast giving steadiness to the one electron, which without it would be so light that it could not be controlled at all. This electron is now to be regarded as a wave of two polarisations and the non-uniform magnetic field has the general character of a doubly refracting medium. As the beam goes through the field it will behave in much the way we saw that an unpolarised beam of light would behave in going through a crystal. It is composed of the two polarised components, inextricably mixed at the start. As it goes, the two components describe slightly different paths, and Fig. 33 (which may be compared with Fig. 29) shows the progress of events. The cross-section of the beam will broaden and one edge will be polarised one way, the other the other, while the middle is still unpolarised. As it goes farther the edges swell out and finally leave a gap in the middle, and the beam has then been completely split into its component parts. At the ends of the beam the non-uniformity is much weaker, and so the two components are not separated there, but still give an unpolarised beam. This is a very good description of what is happening, but it is typical of the relation between the

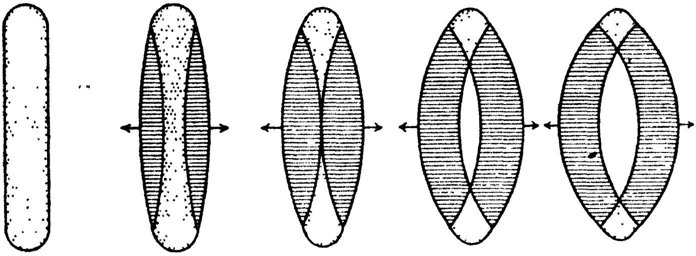


FIG. 33.—SEPARATION OF POLARISED ATOMS

The diagrams show successive stages in the formation of the image of Stern and Gerlach. The dotted parts are unpolarised, and the shaded parts polarised with magnets pointing in the direction of the arrows.

wave and particle aspects of matter that it gives no direct answer to the difficulty propounded above, of how half the atoms jump to pointing along the field and half against it. The answer it makes is to say that the only way of knowing which atoms were pointing either way is to separate them, and this begs the question of how they were pointing before they were separated.

The experiments of Stern and Gerlach show that the atom contains a magnet, but do they show that it is due to the spinning electron? The answer is that the measurements agree entirely with the theory of the spinning electron, and cannot be fitted into any other theory, but that they can never make a hard-and-fast separation between the magnetic effects of the spin and of the orbital motion of the electron. In order to establish the spin we would like to be able to exhibit it for free electrons, in the same sort of way as we can exhibit the polarisation of light. Attempts have been made, or at all events have been proposed, to do this by methods like those of Stern and Gerlach, but closer examination has shown the experiments to be impracticable. At last Bohr gave reasons why we can be certain that no such experiment could work if based on the magnetism of the electron. The ultimate reason is rather subtle, and is based on the principle that we must never mix together the wave and particle aspects of matter. Though we cannot go into his argument here, he showed that if we could directly detect the magnetic force of the spin, then we should be violating this fundamental principle. That we cannot do so depends on the Uncertainty Principle. We must remember that the electron exerts a magnetic force on account of its *motion*, and therefore if we are to observe that due to the *spin*, we must know how to make allowance for the force due to the motion. But on account of the uncertainty of the electron's speed, the force due to the motion must be uncertain to some extent, and the extent turns out to be always greater than the whole effect of the spin.

The magnetism of the free electron provides no help in observing its polarisation. But there are other ways by which it might be observed, though no experiments have yet been successful. The experiment of this type that is easiest to follow is based on theoretical calculations of how an electron is scattered by a nucleus. Mott showed that the scattered wave would be slightly polarised. The only way of testing that it is so, is to submit the electron to a second scattering. He showed that if a stream of electrons is twice scattered through a right angle, the second scattering will be unsymmetrical, so that different numbers of electrons will be found along the lines *BC* and *BD* in Fig. 34. No one has yet succeeded in carrying out this or any other experiment which definitely depends on the polarisation of the free electron. It is still a phenomenon that is certain theoretically, but not verified directly. The proposed experiment exhibits clearly the contrast between the polarisation of electrons and of

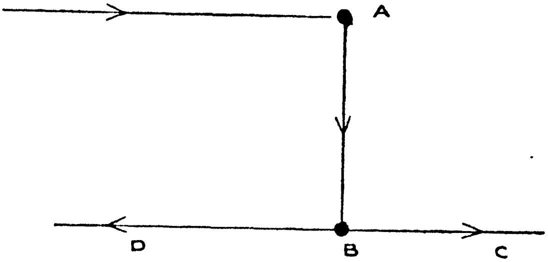


FIG. 34.—DOUBLE SCATTERING OF ELECTRONS

A beam of electrons is scattered by a nucleus at *A*. Some of the scattered electrons go towards *B*, where they meet another nucleus and are scattered again. The numbers going along *BC* and *BD* are compared. It has been proved that *A* will have partly polarised the electrons, and that on this account the numbers scattered along *BC* and *BD* should be different.

light. If a similar experiment is carried out with photons, which is easy to do, the second scattering is equal along *BC* and *BD*, but vanishes for the two directions perpendicular to the plane of the paper.

It remains to consider whether the third of our ultimate things, the proton, also has polarisation. This is a very much harder matter to test. The argument that Dirac used for the electron, based on relativity, applies with equal force to the proton, and with almost complete confidence we assert that it must have spin too. But the same argument indicates that the associated magnetism will be very small, in fact just 1/1845 of that of the electron; we may recall that this number is the ratio of the weight of the electron to that of the proton. The magnetic effects of the spin of the free proton will be masked by the magnetic effects of its motion, in just the way that it was for the electron, for Bohr's argument has equal force in both cases; but there are two main lines of evidence proving the existence of polarisation in the proton.

One depends on certain peculiarities of the hydrogen *molecule*. On account of the magnetism of its two protons, it is found theoretically that the molecule may be constructed in two different ways. We may roughly discriminate between them by supposing that in one the two proton-magnets are pointing the same way and in the others opposite ways; but it is proper to add that this is a misdescription, and that the real distinction is much more subtle. It is found that there is a high degree of stability about the two types, so that a molecule may spend months or years in either state before being knocked into the other by one of the collisions to which gas molecules are subject all the time. In fact, theory indicates that ordinary hydrogen gas should be really a mixture of two gases which only very slowly change into one another under ordinary conditions. This has been verified experimentally. Under the action of extreme cold it is possible to get one type by itself, and if this is allowed to heat up, a type of hydrogen is produced which is abnormal in some respects, and which only returns to the ordinary form in the course of months or years. This surprising fact of chemistry was predicted from one of the subtlest arguments of pure physics, and its verification demonstrates that the proton has the same type of polarisation as the electron.

The other way that experimental evidence can be obtained is by the effects the nucleus produces in disturbing the motion of the electrons in an atom. The magnetism of the protons and electrons stuck together in the nucleus may not quite balance out, and so they may produce a feeble magnetic field. This will disturb the motions of the outer electrons of the atom, slightly altering their energy levels, and above all splitting each energy level into two or more, in something of the manner produced by an external magnetic field in the Zeeman effect. It is observed for several of the elements that the spectrum shows what is called a hyper-fine structure, each of the levels being really composed of several so close together that they can barely be observed to be separate. From the way these levels are arranged, deductions can be made of what sort of magnet the nucleus is. That it is a magnet is clearly shown, and in a general way the theoretical anticipations are borne out; but it is proper to mention that there seem to be in the nucleus some curious things connected with the magnetism which have not yet been explained.

CHAPTER VII

COLLISIONS

HITHERTO we have discussed only questions of a single electron, either by itself, or in fields of force made by other bodies which were supposed to be inert, so that, though they affect the motion of the one studied, they are not disturbed by it in return. This is, of course, an incomplete description. As Newton said, every action must call into play an equal and opposite reaction, and it is only when this reaction happens to be small in its effects that we can legitimately avoid considering it. So now we must discuss how two or more electrons or protons behave when they are treated on equal terms.

We have seen that many of the characters of an electron can be represented by regarding it as a wave-group moving about in space. It would be tempting then to suppose that, when there are two electrons, there are two wave-groups both moving about in space. But this will not do, since two wave-groups are simply a single more complicated wave-group, whereas two electrons are radically different from one. For the correct treatment of the waves of two electrons it is necessary to have two spaces, or in the language used by mathematicians, a space of six dimensions. Of course, there are not really six dimensions, but the mathematician finds it convenient to think and speak in that way. The best picture we can make for ordinary use is something like this. We can make a diagrammatic representation of the wave of an electron by sketching a wave-group on a sheet of paper. It is true this is only in two dimensions, but it is comparatively easy for us to understand three dimensions when we have mastered two. The behaviour of two electrons is described, not by two wave-groups on the paper, but by having two sheets of paper each with one group on it. The second sheet overlies the first. As the motion goes on, each smudge moves about in its own sheet and no transfer from one to the other is allowed. Nevertheless the ink of each group can, so to speak, see what the other one is doing, and experiences a force from it. Since there is repulsion between two electrons, we can say that the patches try to avoid one another. On the other hand, there is attraction between an electron and a proton, and so in that case the two patches tend to come together. This description of the behaviour of two particles applies in cases where they do not approach very closely to one another, but we are going to meet cases where things are not so simple, and where the six dimensions cannot be divided into a pair of threes. I am involved in the difficulty of having to explain ideas which are mathematically fairly easy in terms of the geometry of several dimensions, but which it is by no means so easy to apprehend physically.

It remains to make the interpretation which allows us to understand what happens in ordinary language in terms of particles. The sort of question that may be asked is this. At some instant I take a region of space; what is the chance of finding either both electrons in it or one or neither? Here we, of course, have to think about the ordinary space of experience, and the interpretation is made in a most natural way. We can describe it by imagining the motion of our two overlying sheets of paper suddenly stopped, and we then cut out from them the region that is being searched for electrons. If both sheets happen to be rather black there at that instant, I shall probably have both electrons; if one is black and not the other, I shall have one, and if both are white none. This is only a special case, and I could equally well give the answer to the question, what is the chance that No. 1 is in a region *A* and No. 2 in a region *B*? In mathematical terms we interpret the intensity of the six-dimensional wave as a joint probability for the positions of the two electrons in ordinary three-dimensional space.

We are going to discuss such things as the collision of two free particles, and it will be convenient to consider first how we describe a collision in which one of them is so heavy that it is immovable. The experiment would consist in sending a stream of electrons past a fixed nucleus. In our discussion of the hydrogen atom we likened the electron to a planet going round the sun in an ellipse, but we made no mention of the other possible type of orbit, the hyperbola, which can be described by a comet, and this is what we must now study. Consider first the behaviour of ordinary particles. If the line is known along which the electron-particle approaches the nucleus, its path can be worked out; it shoots past the nucleus, is deflected through a determinate angle, and flies away to infinity with the same speed as that of its approach. But if we have a stream of particles, some will be on one line and some on another, and so they will be deflected to different extents. Several such orbits are shown in Fig. 35. The sort of information that we shall get is that for every thousand incident particles, incident along random lines, say fifty are deflected through angles between 30° and 40°. If the corresponding numbers are counted for various angles, a great deal can be found out about the strength of the force between the nucleus and the electron. It was indeed this method, applied to α-particles, that led Rutherford to the discovery of the charge of the nucleus.

The collision of an electron with a fixed nucleus must also have a wave aspect, but it turns out that the hyperbolic motion has to be expressed in a rather different way from the elliptic. The system is now an open one, in the sense that the electron gets right away again, and for such a system there are no such things as separate normal modes of vibration. We can imagine an experiment in which the electron, say by emerging from a hole, comes as a wave-packet near one of the lines of Fig. 35. Such an electron-wave would behave rather like the hyperbola of a comet, though it would spread all the time, so that the receding branch of the hyperbola would get progressively broader and vaguer. But in the type of experiment that occurs in practice, we have absolutely no knowledge of what line the electron started on, and so we represent its wave aspect by an infinite plane wave. To understand what happens, take the analogy of a set of

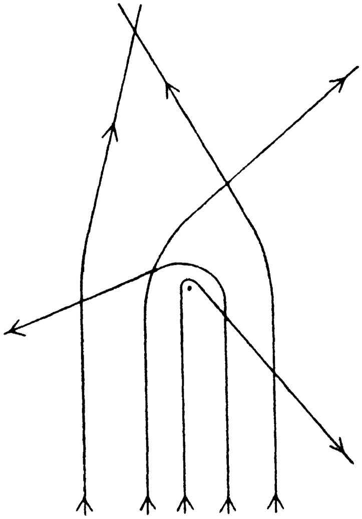


FIG. 35.—HYPERBOLIC ORBITS OF ELECTRON

The electron is deflected by the nucleus through an angle which depends on the distance of the nucleus from the electron's line of approach.

harmonic waves advancing across the sea. At one place they encounter a small obstacle, say the fixed round post shown in Fig. 36. This will disturb the wave system, with the result that there will be superposed a set of circular waves, travelling out from the post, of the same wave-length as the incident waves. The same sort of thing will happen for the electron and nucleus; there will be a spherical wave, of the same wave-length as the incident, travelling out from the nucleus. We may describe the relation of wave and particle by saying that the wave does not know how near its particle went to the nucleus, and so cannot tell what sort of a hyperbola it is describing, and therefore the wave must do the best it can by offering all possible alternatives. The detailed solution yields the intensity of the spherical wave in each direction, and that is interpreted as giving the probability of the particle being observed, say by means of a scintillation, in the

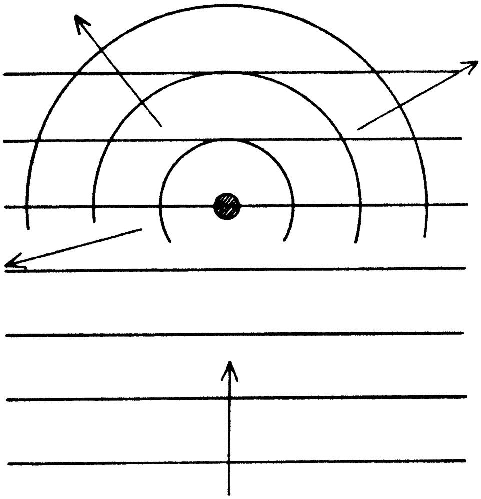


FIG. 36.—WAVE SCATTERED BY SMALL OBSTACLE

The fixed obstacle scatters the incident plane wave and gives rise to a spherical wave travelling outwards in all directions from the obstacle.

corresponding direction. The point to be emphasised is that the wave does not tell us what happens, but tells us everything that *may* happen with the appropriate probability.

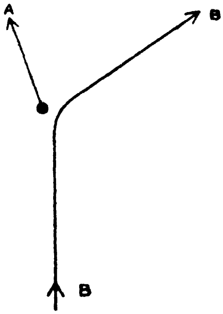


FIG. 37
COLLISION BETWEEN TWO FREE PARTICLES
A is at rest and is struck by *B*. Their subsequent motions are controlled by the conservation of momentum and energy.

We will now take a typical case of the interaction of two particles, considering first the particle aspect, and then afterwards examining how the same results come from the wave aspect. We will take what may be called the billiard problem. This is illustrated in Fig. 37. A particle *A* is lying at rest, and another *B* is projected with given speed so as to collide with it. For billiard balls the collision takes the form of a sharp impulse as the balls touch, but we need make no restriction of this kind. We need merely suppose that the particles exert no forces on one another except when they are fairly close together; when they are near, their mutual force on one another may be either attractive or repulsive. In consequence of the collision, *B* is deflected from its original line and *A* is knocked away. The directions in which they go, as every billiard player knows, will depend on how they collide, but the theorems of conservation of momentum and energy impose great limitations on the possibilities. For example, there can be no momentum after the collision in the direction perpendicular to the line of *B*'s approach, because there was none before. It follows that if *A* goes to the left, *B* must go to the right, and when we know how fast *A* is moving we can deduce the speed of *B*. Working out the conditions in detail, it turns out that if the line of *A* after the collision is known, then from it we can deduce the line of motion of *B*, and the speeds of both *A* and *B*.

This is true for any collision of two bodies, but we have to consider that we cannot control the collisions of electrons, as the billiard player controls the line along which his ball approaches the other. All we can do is to fire a group of *B*-particles at a group of *A*-particles and see the result. The consequence is that all sorts of different collisions will take place, the *A's* being sometimes projected one way and sometimes another; but in every case the *B's* will travel off in the direction determined by the conservation theorems. We may describe the process by saying that the billiard player is blindfolded and compelled to repeat the same stroke again and again, with the balls replaced each time, and a count is kept of how the balls behave each time. To make the experiment more definite, let us suppose that the observation is made by setting a target in some place, which shows a scintillation whenever an *A* strikes it. The observer would then get the information that for every thousand *B's* fired, let us say ten *A's* will hit the target. He would express this in the form that there is a one per cent. chance of collisions giving an *A* going towards his target. But now let us elaborate the experiment, by supposing that there is a second observer who has a target that can scintillate when it is struck by a *B*. If he sets it up anywhere and counts scintillations he will know what is the chance of a collision throwing a *B* towards him. But we have seen that associated with each *B* there is an *A*, and he may happen to have set his target in such a place that it gets exactly the *B's* which have recoiled from the *A's* observed on the first target. In this case the observers at the two screens will always observe scintillations simultaneously. Experiments of this kind have been done for various types of particle, and we shall come to them again later. Looking at the matter crudely we cannot imagine anything more typical of particles than this simultaneity in the scintillations.

We must now turn our attention to the wave aspect of the same question. If we looked at the collision superficially we should describe its wave aspect in some such way as this. We have an incident plane *B*-wave, falling on a wave-packet of *A* more or less at rest—no wave-packet can be absolutely at rest on account of the Uncertainty Principle. Then *B* is scattered into a spherical wave, and so is *A*. This might seem to be the natural extension of the case where *A* was fixed; but if it were the right extension, what would become of the experiment where we observed simultaneous scintillations of *A* and *B* in two definite places? It would seem that we should expect to get *A's* and *B's* anywhere and without correlation. This is because the argument has been wrong; there are not two spherical waves of *A* and *B*, but a single wave of *A* and *B* in six dimensions.

It is not hard to consider such a wave mathematically, but I have not succeeded in finding a simple geometrical analogy which will explain how the six-dimensional wave is to be understood as a description of the motions of the two particles. The best that can be done is the rather imperfect analogy sketched in Fig. 38. Consider waves advancing across the sea, and make the convention that the north and south directions belong to *A* and the east and west to *B*. Then if we measure the distance between crests northwards, we get the wave-length belonging to *A*, and interpret it as its momentum in the particle aspect, while the wave-length measured eastwards will give the momentum of *B*. If this wave meets an obstacle, say a post stuck up in the sea, it will give rise to a circular wave going out in all directions. The frequency of this scattered wave is the same as that of the incident, and since in the particle aspect frequency becomes energy, this means that energy is conserved in the collision. Now consider the wave scattered in some direction. If we measure the distance between crests northwards and eastwards we shall evidently get a pair of values differing from those of the incident wave, and in the particle aspect this implies that there has been an interchange of momentum between *A* and *B*. Moreover, by taking any other direction of the scattered wave we get a new pair of momenta for *A* and *B*. In the wave aspect all possible pairs of momenta

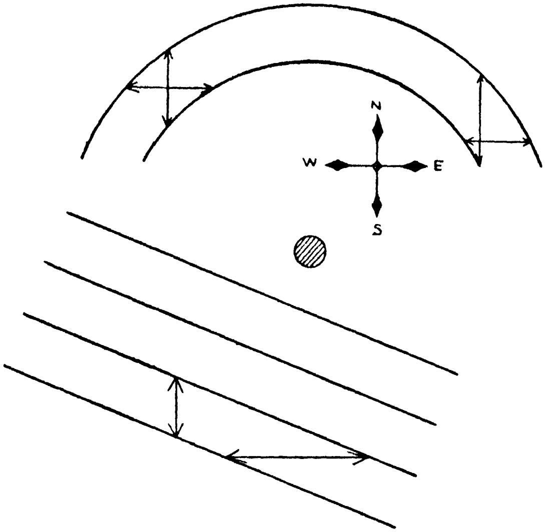


FIG. 38.—WAVE OF TWO PARTICLES

An analogy which illustrates how the wave aspect of two colliding particles must be treated. Wave-length is usually measured perpendicular to the lines of the crests, but we now depart from this rule. We measure the distance between crests northwards and call it *A's* wave-length, and eastwards and call it *B's*. These wave-lengths are interpreted as momenta in the usual way. When the wave is scattered, though the wave-length in the ordinary sense is unaffected, yet *A's* and *B's* wave-lengths are changed. This corresponds to the interchange of momentum in a collision.

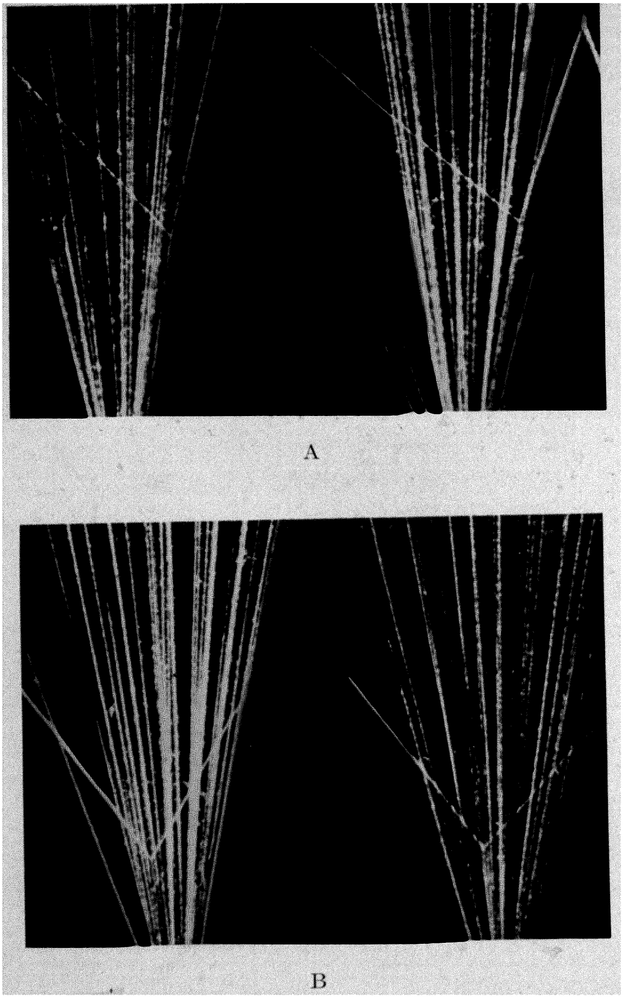
are offered to us simultaneously, just as in the earlier problem of an electron going past a fixed nucleus all possible collisions were offered at once. The wave again expresses not what will happen, but everything that might happen. It does not say "*A* will go here and *B* will go there," but "*A* may go anywhere, and I can tell you what is the chance of its doing so; but if it does come here then *B* will certainly go there."

The study of the collision of *A* and *B* is very illuminating in attaining an understanding of the limitations of the half-worlds of the particle aspect and the wave aspect. If we only think of the particle aspect, then we cannot understand a great many things like diffraction and the way that atoms are put together. But if we try to pay attention only to the wave aspect the trouble is of quite a different kind. The trouble is that nothing ever happens. The waves are like the expert advisers of a government, who tell the ministers what will be the consequences of every proposed policy, but, as is perhaps not unknown even in the civil service, refrain from ever deciding which policy should be chosen; it requires the decision of the minister to change the deliberation into action. The half-world of waves is very wise in that it foresees everything, but it is a dead world because nothing happens until it is animated by an observation drawn from the particle world. As soon as an observation has been made, the event comes to life, and all the possibilities that did not happen may be wiped out.

The new physical theory differs from the old in that observation plays an active part in it, instead of being merely taken for granted; and we have to convince ourselves that this will not disturb the fundamental principles of knowledge. We can see how it comes about that it is meaningless ever to ask what happens, unless we have an observation of some kind to tell us. Take, for example, the case where we have the apparatus for making an experiment, but do not make it. Thus an observer is watching for scintillations of the particle *A*, but for the moment his attention wanders. Our instinct is to say that whether he was looking or not cannot really make a difference. But we are in a dilemma. If the experiment has no visible collateral effects whatever, then we have no means of verifying whether it was unaffected by his lack of attention. On the other hand, there may be associated effects, such as the simultaneous scintillation of *B* seen by the second observer. If this man is more attentive and sees his scintillation, then we can confidently say that there was one made by the *A* as well, even though the observer missed it; but the point of our question has disappeared, because we know that there was one there, through the *observation* of the associated effect of *B*. This is a typical example of one of the most satisfactory features of the new philosophical outlook. It is meaningless ever to ask what has happened, unless there is either a direct or an indirect observation to tell one; and if there is an observation it is pointless to ask, because one knows the answer. The new mechanics represent the perfect antithesis to mysticism, for it denies sense to any question unless we can verify the answer by an observation.

We will now turn our attention to some specially interesting types of collision. In view of the difficulty in finding physical models of the wave aspect which can be visualised, I shall only make rather brief descriptions of some salient points in each. In Plate I, A we have the type of collision of the billiard problem. The moving *B* is an α-particle, which we may recall is the nucleus of a helium atom; the *A* is an oxygen atom, a good deal

PLATE V



ALPHA-RAY TRACKS

Two photographs were taken from different directions, so that the tracks could be reconstructed in three dimensions. (*Photographs by Mr. P Blackett*)

A. An α-particle collides with a hydrogen nucleus. The hydrogen flies away to the left, and the heavier α-particle is a little deflected to the right.

B. An α-particle collides with a helium nucleus. The angle of the fork is a right angle, and it is impossible to say which is the nucleus and which the recoiling α-particle.

heavier. Blackett has taken photographs involving more than a million α-ray tracks, and has thus got a few with peculiarities of many and various kinds. If the gas through which the α-particles pass is hydrogen, the α-particle *B* is four times as heavy as the hydrogen *A*, and consequently may project it at a speed higher than its own. This is shown in Plate V, A. If the gas is helium, then *A* and *B* are identical. It can be shown by the principles of conservation that they always diverge from one another at right angles, and this is confirmed in Plate V, B. In this case there is absolutely no way of knowing which of the two particles after the collision is *A* and which is *B*. Moreover, there is a most interesting theory of the collision of *identical* particles, quite unforeseen until it was recently worked out by Mott, which has a marked effect on the comparative frequencies of occurrence of different angles after the collision. We shall discuss this at length in the next chapter. Here it may be mentioned that there are twice as many 45° collisions as one would have expected, and this is compensated by a greater rarity in the numbers at other angles.

The next collision is also connected with α-particles. We have been speaking of the α-particle as a wave, whereas the photographs showing the tracks strongly suggest that each track is like the path of a bullet, and not a bit like a wave. The matter is even more surprising if we consider the origin of the α-particle in the radium. A very beautiful theory of this has been made by Gamow from the wave aspect. The α-particle is a wave confined in a spherical box in the nucleus with walls that are nearly perfect reflectors. It oscillates to and fro with spherical symmetry, but there is a slight lack in the perfection of the reflection of the walls, so that a little of the wave is all the time leaking out. The wave aspect then represents the α-particle as being a continuous very feeble wave given out by the atom, and it shows the extraordinary difference between the wave and the particle aspects when we think that radium lasts for some thousands of years, and is all the time giving out the spherical wave steadily in all directions. How are we to reconcile this with the obvious suddenness and asymmetry of the track of the α-particle? The answer is found to be quite satisfactory from the wave aspect; that it seems unnatural is due to our inability to picture the many-dimensional space. It can be shown on pure-wave principles that the spherical wave has a certain chance of ionising an atom—that is, of putting it into the condition of producing the water drop which constitutes an element of the track; and there is a probability practically amounting to certainty that any second atom ionised will be on the same line and that the ionisation will occur just after the first.

The third type of collision which we consider is called the Ramsauer effect. This was discovered in an investigation of how electrons are scattered by atoms. This question has been much studied, and we may recall that it was in examining the scattering of electrons by nickel that Davisson first found that he was investigating not the scattering by atoms, but by a crystal. In a gas, however, there is no systematic arrangement of atoms, and so whatever is observed is due to individual atoms. Electrons going at an enormous speed, like the β-particles of radium, go in straight lines right through everything, with occasional sharp deviations where they get too near a nucleus or another electron. Slower electrons are much more easily deflected, so that a beam travelling through a gas is gradually dispersed because more and more of the component electrons wander off the original line.

Ramsauer discovered the surprising fact that under certain circumstances electrons going at a very slow speed are much less dispersed than swifter ones. On the old particle theory this seemed almost impossible to explain; the scattering at higher speeds showed that there were forces between electron and atom; how could they avoid affecting the electron at low speeds? To put it in picturesque language, the atom can only know that the electron is going slowly, and therefore that it must be left alone, by feeling its speed; it can only feel its speed by the forces of interaction, but by the time these forces have had enough play to give the information asked, the harm is done and the electron deflected. It is found that this difficulty disappears in the wave theory. To use the same picturesque language, in that theory everything knows about everything else all the time, so that the atom is not taken by surprise in the same way.

We will next turn our attention to the interaction between photons and electrons or atoms. There is considerably more difficulty in describing what happens in these collisions than in those of electrons, because the present theory of the interaction of light with matter is really rather unsatisfactory. The theory, in its final form due to Heisenberg and Pauli, is extremely difficult; indeed it has been only in the hands of a few of the leading workers in this field that anything has been made of it, and it is, I think, rather widely felt that it is not founded on the right lines. Apart from this theory there exist approximative methods which appear to be satisfactory, but they are avowedly approximate and no one quite knows what they are approximating to. The principle on which the interaction of photons and electrons has been worked so far is roughly this. First the photon *B* is regarded as so slightly influenced by the other particle *A*, that *B's* effect on *A* can be found without paying attention to the reaction of *A* on *B*. Then when the influence on *A* is known, it is found that *A* would send out light; the amount is calculated and can be regarded as the scattered part of the photon *B*. The justification of the process lies in the fact that the interaction is rather slight; it is perhaps sufficient for most purposes, but it inevitably fails to answer certain interesting and quite fundamental questions. However, this is the best method that we have at present, and as I have said it gives a fairly good account of many things. But in describing them it will be inevitable that we should mix together the wave and particle aspects in a way which was not necessary for collisions of electrons and which I do not believe to be ultimately correct.

The first case which we shall take is the Compton effect, to which we have already referred.[1] In some ways this is the most primitive type of collision that there is, for it is the interaction between one photon and one electron. Compton's original discovery of the effect was due to consideration of the particle aspect of the photon at a time when it was a far more revolutionary idea than it is now. The electron is supposed at rest, and the photon collides with it. The photon has a definite momentum and energy, which in the wave aspect signifies that it has a given wave-length and frequency. It only differs from a material particle in that the speed of motion of a photon is always the speed of light and has nothing to do with its momentum. The collision is an example of the billiard problem, and is governed by the conservation of momentum and energy. From these principles it is possible to calculate all the characteristics of the motion after the collision, provided that the direction of motion of *A* is given. But again we get no knowledge of what is the direction of motion of *A*, because we cannot control how the collision will take place, but must merely study statistically the effects of a large number of collisions. Since the momentum of the photon *B* is known when its direction is given, its wave-length is also known. Its momentum will always be less than that which it had when approaching *A*, since it has to give momentum to *A*, and so the scattered wave-length is longer than the incident. For a slight collision in which the photon is only scattered through a small angle the wave-length is only lengthened a little, but when the collision is head on, so that the photon is scattered straight back, it attains its maximum value. It turns out that there is a very simple rule to determine the change of wave-length; when the photon is reflected straight back its wave-length is lengthened by $4{\cdot}8 \times 10^{-10}$ cm. This is a very small amount; and in the scale model where we took the atom about a yard in radius, this length would be about two inches. It is comparatively easy to detect a change of two inches in the length of a yard and very difficult in a mile, so that the experimental verification must be done with the short waves of X-rays, and is hardly possible with visible light. Compton verified this change of wave-length at various angles of scattering for certain wave-lengths. Later he and others, with the use of more elaborate apparatus, detected the recoil of the electron, and by observing both electron and photon simultaneously showed that the conditions of momentum and energy were satisfied. The electron acquires velocities of the order of a tenth of the speed of light in recoiling from X-rays, and is competent to produce a short cloud track of the kind we have already seen for α-particles.[1]

---

[1] See p. 95.

---

[1] The photographs are very hard to decipher, and so none are reproduced here.

These results were quite revolutionary when they were discovered, but have fitted very well into the scheme of the new quantum theory. As far as they go, the experiments I have described supply no new knowledge beyond the verification that the photon obeys the conservation laws. But we have not yet considered one part of the matter. The *association* of *A's* and *B's* motions was verified, but what will be their chances of performing any particular pair of motions. For example there will be a certain number of photons scattered straight back, a certain number at right angles, and a certain number barely deflected. We know the momentum and wave-length in each case, but it is also important to know the relative numbers for the various directions. The importance of this question is due to the fact that it will show what is the nature of the interacting forces between electron and photon; the previous work merely verified the conservation theorems. The first experiments on the subject were made by J. J. Thomson, with X-rays of rather long wave-length, long before the Compton effect was discovered. Long wave-length means little momentum, and in these experiments the recoil of the electron was negligible. In that case the scattering could be worked out by the use of the old classical theory, and it gave a simple rule. The radiation is equal in front and behind, and twice as great as it is sideways; the dependence on the direction only arises because of the polarisation of the X-rays. For shorter wave-lengths this simplicity is gradually lost, and both the experiments and the theory become much more difficult. Into this it is not possible to enter, but it may be said that a satisfactory agreement is reached only when the theoretical calculations make allowance not only for the polarisation of the light but also for that of the electron.

The collision of a photon and a free electron is the most primitive type that can occur, and it may seem surprising that it should be so complicated. The reason is partly that this collision is so primitive that we can study all its details. In the other interactions that we shall consider, this is not so to quite the same extent, and consequently they are apparently simpler. The simplest case of all is the photo-electric effect. When light of a definite wave-length falls on a suitable metallic surface, it is found that electrons come off from the surface at a definite speed. The electrons are moving about freely between the atoms of metal in a way too complicated to analyse in detail, but the conservation of energy gives all the necessary information. The incident photon has definite energy, and if it is absorbed this must appear elsewhere. It is acquired by one of the electrons and makes possible its emergence from the surface of the metal. This explanation of the photo-electric effect was due to Einstein, and it is notable as one of the very earliest positive developments of the quantum theory.

We next consider the collision of a photon with an atom, and it will suffice to take a hydrogen atom consisting of a single electron tied in its lowest energy level to a heavy nucleus. A number of different things can happen. If the photon has very short wave-length like X-rays, the forces it exerts on the electron are far greater than those of the nucleus, and there will simply be a Compton effect, which alternatively may be regarded as a photo-electric effect. If the wave-length is rather longer, the forces are more equal, and there will appear certain departures from the Compton effect, and a transitional state of affairs which we need not consider. The question that concerns us now is how a photon of visible light behaves when it strikes an atom. To consider this we must remember the characteristics of the radiation of the atom. We recall that it has certain energy levels, normal modes in the wave aspect, and that it radiates light by dropping from a higher level to a lower.[1] The light radiated has energy equal to the difference between the energies of the two associated levels, and hence it follows that the frequency of the emitted light is equal to the difference of the frequencies of the two associated normal modes.

From the particle aspect the description of emission consists in the drop in level and the consequent creation of a photon. The converse process to emission is absorption, and so absorption must be described as the disappearance of a photon in the atom and the consequent rise of level of the electron in the atom. In the wave aspect these phenomena are not described in a simple manner, since the technical description of the creation or annihilation of a photon-wave is very troublesome. But without attempting to enter into these difficult questions we can understand certain properties of the process of absorption which are interesting. We are to regard the light as a plane-wave falling on the atom, and the fact that it is absorbed shows that beyond the atom the intensity of the light must be reduced. This reduction occurs by the act of the atom, which must therefore be supposed to *emit* a wave, but a wave of such a phase that it interferes with the original wave and tends to reduce its amplitude. Now it is not possible for a wave from an atom to go in a single direction; it must go out all round, and in the other directions there are no waves for it to interfere with. So we deduce that the process of absorption is inevitably associated with a strong scattering of light. This argument explains a well-known property of matter which appears paradoxical at first sight. Substances such as metals which are very opaque to light are very good reflectors; for opacity means strong absorption, and good reflection means strong scattering.

---

[1] See p. 122.

The phenomenon called *resonance radiation* is another example of the same thing. This radiation can be excited in many vapours such as mercury. Light from a mercury arc falls on an evacuated bulb containing a little mercury vapour. The light that is operative happens to be ultraviolet, so that nothing is to be seen by the eye, but with appropriate devices it can be shown that the bulb scatters the light from the arc strongly all round. The atoms of mercury in the arc are raised from level 1 to level 2 by the electric current, and they then fall back to level 1 with emission of light 2⟶1. This light falls on the bulb, where it finds atoms in level 1. It raises them to level 2, with the absorption of the light, and they then fall back to level 1 re-emitting this light all round. That is the description in the particle aspect. In the wave aspect the atoms in the bulb emit a wave of the same frequency as was incident on them, but in such phase that they weaken the intensity of the light that has passed them, and this wave, being necessarily spherical because it has emerged from an atom, implies a scattering into the other directions.

Our last example of a collision will be the case where the frequency of the photon is not one of those which can be absorbed by the atom. This is the most familiar of all in nature, since it corresponds to the ordinary refraction of light by transparent substances. Though the process has been fairly completely mastered, it is not very easy to explain in simple terms. Fig. 39 shows roughly what happens. The incident photon tries to raise the level of the atom to a level X, which is determined by adding on the photon's energy to the energy of level 1. The atom, however, has no natural level at X, and so falls back *at once* to the original level 1. In doing so, it emits light of the same frequency as the incident light, but this time it turns out that the phase is not like that in absorption, and there is practically no diminution in the intensity of the transmitted light. There is, however, a change in its phase, and this can be made to explain the ordinary refraction of light. Moreover, there is inevitably some scattering of the light, and we are very familiar with one example of this in the light of the sky. The amplitude of the scattered wave can be deduced by comparing the brilliance of the sky with that of the sun; while by the study of the refraction of the air another measure of the amplitude can be made, which depends on the interference between the scattered wave and the incident. These must, of course, agree, but in order to make them do so, it is necessary to know how many atoms there are in each cubic centimetre of air. It was by the use of this method that Rayleigh made

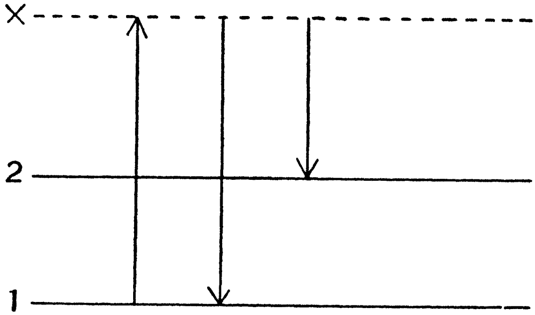


FIG. 39.—PHOTON SCATTERED BY ATOM

The photon raises the energy of the atom to level X, but as this is not one of the atom's energy levels, the atom falls back at once, scattering the light. If it falls back to level 1 we have ordinary refraction; but if to level 2 then we have the Raman effect.

the earliest reasonably good determination of that fundamental physical number, the number of atoms in a cubic centimetre of air.

The same theory which explains ordinary refraction also predicts another effect. For ordinary refraction the photon rises to a non-existent level and then falls back. But it is possible that it should not fall back to the old level; it might rise from 1 to X and then fall from X to 2. In this case the light scattered is different in frequency from the incident; it is given by adding to or subtracting from the incident frequency an amount equal to the frequency 2→1. This radiation, being of a changed frequency, cannot interfere with the incident light, and so is only observed as scattered light, and has no effect on the refraction. The effect was predicted theoretically some years ago, but was only discovered experimentally by Raman about three years ago. Raman sends the light of a mercury lamp on to solids or liquids, and measures the frequency of the scattered light. The solid scatters light firstly of unchanged frequency, but this is accompanied by other light of a little lower and also of a little higher frequency. The new frequencies indicate the presence of a pair of levels rather close together, and by a study of the Raman effect it is thus possible to map out the energy levels of a solid. Before the discovery of the Raman effect this was an exceedingly difficult business, since it involved work in the remote infra-red. Now it is possible to explore this infra-red region making use of visible or ultra-violet light, and this is technically a much easier process. It is a truly remarkable development of the new physics, that experiments making use only of ultra-violet light should provide the most powerful line of attack on phenomena in the infra-red, and so indirectly on the nature of the solid state.

CHAPTER VIII

THE EXCLUSION PRINCIPLE

THERE exists a principle in the new quantum theory which pervades the whole subject to an extraordinary degree, and yet which occupies a curiously isolated position in it. It is called the Exclusion Principle of Pauli, who was its discoverer. The rest of the principles of the new mechanics all hang together and belong to a single realm of thought, and they seem as though they could have got along fairly well without the Exclusion Principle, though, of course, the world would have been a very different place if it had not existed. There are two chief ideas in it. One is that electrons can never be distinguished from one another in any way, and the other is, roughly speaking, that electrons avoid one another.

Though the rest of the theory would have made a consistent whole without the Exclusion Principle, we can find a point of contact between them. We have seen in other parts of the theory how systematically Nature refuses to allow a meaning to any question we ask, unless we can answer it ourselves. Now since all electrons are absolutely alike, we can never devise an experiment which will enable us to recognise a particular one, and the principle therefore insists that we ought not to talk about electrons as having any individuality. Consider a collision between two electrons, as in Fig. 40. *A* is at rest and is struck by *B*. Both diverge from the collision in directions at right angles to one another, and as we cannot paint *A* red and *B* blue, there is no way of knowing which of them went to the right and which to the left. It follows that we may not even ask which was which, but may only say that two went into collision and two came out, and must not raise the question of their individuality at all. The Exclusion Principle embodies these conditions, but it is at present expressed in a way that will at some future date probably appear rather clumsy. The description consists in starting as though the electrons had full individuality, and then showing that this individuality does not matter. To understand it we must first consider certain peculiarities of a system containing identically similar parts; these are called the exchange phenomena.

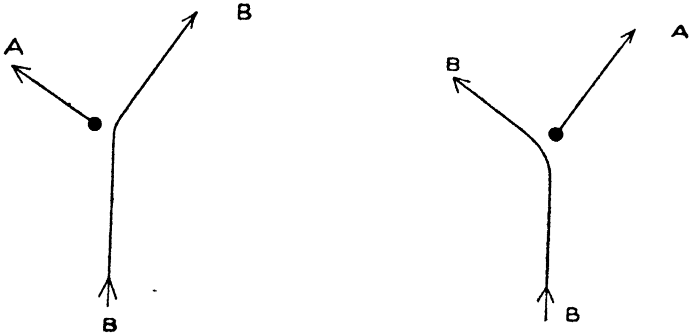


FIG. 40.—COLLISION BETWEEN IDENTICAL PARTICLES

A is at rest and is struck by *B*. On account of their equality of mass they always diverge at right angles to each other, and there is no way of recognising which is *A* after the collision.

Everyone knows the disagreeable jangling sound that is made by a note of a piano that is out of tune. Two of the wires which ought to be helping one another in producing the same note have got out of tune with one another, and produce what are called *beats*. Their frequencies of vibration are slightly different, so that the waves they send out alternately agree and disagree in phase, and so the note is alternately loud and soft. When the wires are only a little out of tune the beats are slow and can be heard one by one as a throbbing sound, but in worse cases they may be so rapid that they merely produce a disagreeable jangle. The beats depend on the strings being a little different from one another in tension, but a similar phenomenon can occur in a system composed of two identical parts.

Suppose that we have two pendulums identical in every respect, but mounted close together on a support which is not absolutely unyielding. We set pendulum *A* in motion, while *B* is at rest. If the support were quite rigid *A* would go on swinging, and *B* would stay at rest. But there is a small effect on *B* through the give of the support, and consequently *B* begins to move. What happens is rather striking, for *B* starts swinging more and more, and at the same time *A's* motion diminishes until *B* is swinging to the full amount while *A* has come to rest. Afterwards the motion is transferred back to *A* again, and it continues passing from one to the other until the motion has died away altogether.

The reason for this behaviour is easily understood. The system of the two pendulums, like every other vibrating system, has normal modes of vibration, but these do not consist one in the motion of *A* and the other in the motion of *B*. The modes are shown in Fig. 41. In one of them *A* and *B* are swinging equally both to the left at the same time, and in the other they are also swinging equally, but now when one is to the left the other is to the right. To distinguish the two modes I shall call them by names of which the application is not very obvious for the pendulums, but which are convenient because they are used in the quantum theory. The first is the symmetric mode, the second the antisymmetric. The two modes have frequencies of vibration which differ from one another a little, because the effect of the yielding support is different in the two cases. The motion we described before in which *B* and *A* alternately come to rest is the superposition of these two vibrations. At one time the phases of the two modes are such that *A* is to the left at the same time for both. Consequently *B* would be to the left for one and to the right for the other, and so by the principle of superposition it is at rest. Later, on account of the slight difference of the frequencies of the modes, it will occur that the two modes are in opposite phase for *A*, and therefore in the same phase for *B*, and consequently the motion will now have passed entirely to *B*. The exchange of the motions is illustrated in Fig. 42. It is perhaps well to notice that in likening the exchange of motion of the pendulums to the beats of the two piano wires, the analogy is not between each wire and each pendulum, but between each wire and each mode of the two pendulums. It is the two modes that beat together.

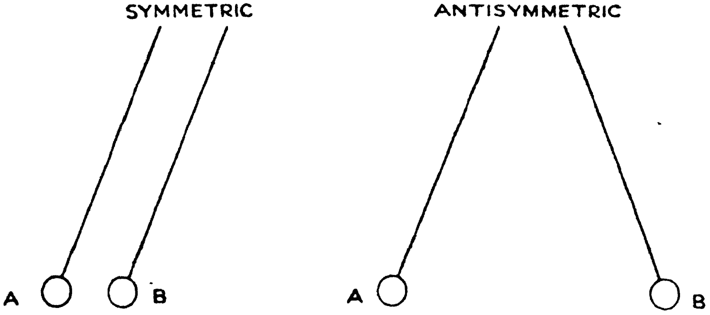


FIG. 41.—NORMAL MODES OF TWO EQUAL PENDULUMS

If the pendulums were quite independent the modes could be taken as, firstly, the motion of *A* without *B*, and, secondly, that of *B* without *A*. When they are linked, however loosely, the normal modes are as shown. *A* is supposed a little in front of *B*, so that they never hit.

The example of the two pendulums illustrates a principle that is universal. When two identical vibrating systems are linked together, the normal modes consist of two types, symmetric and antisymmetric, which have slightly different frequencies. In the symmetric modes the two parts are simultaneously in the same phase, in the antisymmetric in opposite phases. If one of the two is started while the other is at rest, beats will occur; the motion will be handed to and fro between the systems. The beats are slow when the frequencies of the symmetric and antisymmetric modes are nearly equal, and this will be the case when the linkage of the two systems is weak, so that their mutual influence is small. In cases of strong linkage the frequencies of the modes may be very different, and the beats will be rapid, perhaps so rapid that they can hardly be recognised as such, but the characters of symmetry and antisymmetry still hold.

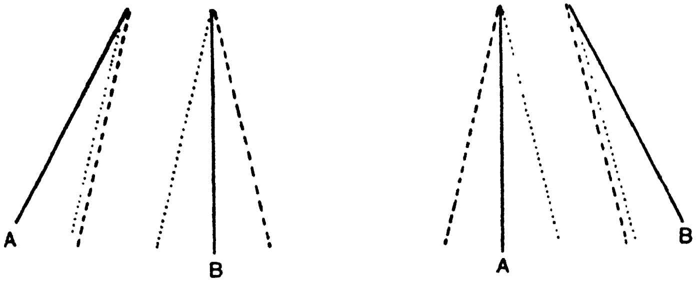


FIG. 42.—EXCHANGE OF MOTION BETWEEN PENDULUMS

The actual motion is a superposition of the two normal modes. The symmetric mode is shown dotted, the antisymmetric mode dashed, and their resultant is represented by firm lines. On the left, both modes reinforce each other for *A* and cancel for *B*, so that only *A* swings. On account of the slightly different frequencies of the modes, the system will later reach the state shown on the right, where *B* swings and *A* is at rest.

We will apply this principle to two electrons in an atom, introducing the exchange phenomenon, but at first still supposing that we can distinguish between the two electrons. Later we shall materially modify the model in removing this condition. The nucleus of the helium atom contains four protons and two electrons and therefore has net charge 2, and so it can bind two other electrons outside. We will imagine that these are brought up in turn towards the nucleus and consider an example of how they could be placed. The first finds the nucleus quite bare, and can therefore be bound in any of the modes of the type we described for hydrogen. The only difference will be that as the central force is stronger the corresponding mode will here be closer in. We will say that it is bound in the mode that would correspond to the first circular orbit. The second electron is now to be bound, and we will suppose that it is bound rather far out also in a circular orbit. At places far out the force due to the first electron will be nearly as though it came from the nucleus, so that effectively there will be a nucleus of charge 1, and the circular orbit will be in the same place as it would be for hydrogen. The orbits are sketched in Fig. 43. For simplicity they have been drawn as lines, but really they should be hazy rings like Fig. 25. Moreover, it must be remembered that though we draw the modes on the same plane, they are really to be imagined on different sheets of paper, one overlying the other. The linkage between the two electrons is weak,

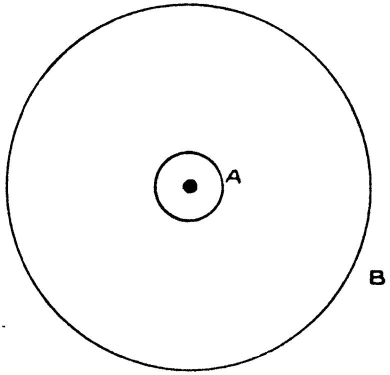


FIG. 43.—TWO ELECTRONS IN AN ATOM

The circles show the orbits of the electrons *A* and *B*; they should really be hazy rings. The beats of the two modes are here represented by the electrons periodically exchanging places.

since the inner one is so near to the nucleus that the outer one exerts only a very weak force by comparison, while for the outer one the force of the inner one is always nearly the same as though it were at the nucleus.

We must remember that though the greatest intensity of the modes is at the places shown in Fig. 48, yet in fact each mode does have some slight intensity all over space, so that, for example, if we looked for electron *A* with a microscope we are most likely to find it near the circle marked *A*, but there is a small chance of seeing it anywhere else. The point is that in the wave theory, unlike the case of ordinary mechanics, there are no strict barriers, but only probabilities, to prevent the electron from being found in unexpected places. We now have to make allowance for the exchange phenomenon. This is here represented by a periodic interchange of orbits between *A* and *B*, which is the analogue to the handing back and forth of the motion of the two pendulums. The mere fact that there is an interaction between the two electrons makes it a matter of dynamical necessity that they should periodically exchange places. The linkage is weak here, and therefore the exchange will not take place often. To summarise what we have got, we are supposing that we can recognise electron *A* when we see it. Then we may find it anywhere, but at first there will be a strong probability of its being somewhere in the inner circle, half a beat later an equally strong probability of its being in the outer circle, at the next half-beat back in the inner, and so on. The opposite will be true of electron *B*.

In describing the two pendulums we saw that there was an alternative description in terms of the two normal modes. The same is true here, but the physical motion corresponding to either of the separate modes is not at all easy to visualise. We saw in the case of hydrogen that when an electron-wave is in a normal mode, the electron-particle is equally likely to be found anywhere round the circle, and this we may express by saying the electron is all over its orbit all the time. The same will be true here. In either of the two normal modes both *A* and *B* will be all over their orbits all the time, and the orbits now each of them include both the circles. For a single mode there are not beats, so that if we look for *A* at any time we are equally likely to find it on the outer or inner circle. The distinction between the two modes is very easily expressed mathematically, but is not very easy to visualise. The best description is this. Suppose that we look for both electrons at the same instant by means of a flashlight X-ray photograph. Then, though we may find the electrons anywhere, we are much more likely to find one of them on the inner circle, and the other on the outer. The question of symmetry or anti-symmetry will be revealed by where they tend to lie on the two circles, but for most pairs of circles, the comparative differences of probability are very slight.

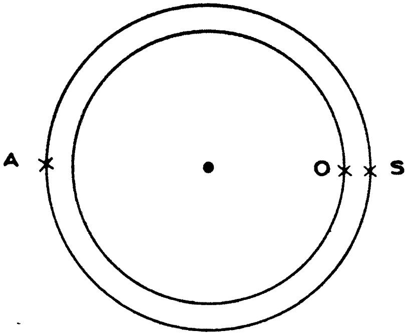


FIG. 44.—SYMMETRY AND ANTISYMMETRY OF ELECTRONS

Two electrons are in orbits with angular momenta differing by one quantum. If one is observed to be at *O*, it is most likely that the other will be found at *S* in the symmetric mode, but at *A* in the antisymmetric mode.

The distinction is best seen by taking the pair of modes corresponding to two adjacent circles, instead of having one close to the nucleus and the other far out. By adjacent circles are meant those for which the angular momentum differs by a single quantum. This case is shown in Fig. 44. If one electron is found to be at *O*, then in the symmetric mode the other is most likely to be at *S*, whereas in the antisymmetric mode it is most likely to be at *A*. We can thus roughly characterise the two types by saying that in the symmetric mode the electrons tend to come together, whereas in the antisymmetric mode they tend to avoid one another. That this is only a very rough statement is illustrated by taking the case where the angular momenta of the orbits differ by two quanta; here the probabilities are indicated by Fig. 45. Mention must also be made of one important exceptional case. If we try to make both electrons go into the same orbit, it turns out that there is no antisymmetric mode at all, but only a symmetric, as is indicated in Fig. 46.

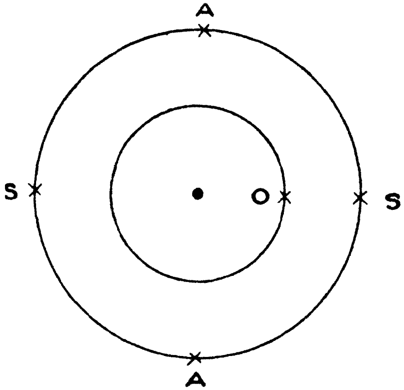


FIG. 45.—SYMMETRY AND ANTISYMMETRY OF ELECTRONS

Two electrons are in orbits with angular momenta differing by two quanta. One is observed to be at *O*. The most likely places for the other are at *S* if the mode is symmetric and at *A* if it is antisymmetric.

We have spoken of the two pendulums as being identical systems, but in one respect they are not so, for however alike they may be in all other ways, they cannot be in the same place. The two electrons, on the other hand, have not this difference, for we have just seen how their motions will entirely overlie one another, and this fact leads to a very important consequence. A force which acts on one electron will act to an equal extent on the other, and this no matter how violent the force may be. It can be shown that if the two electrons start in a symmetric mode any such disturbing force may alter their mode of motion, but cannot give rise to any vibration in an antisymmetric mode. Similarly if the atom starts in an antisymmetric mode, the external forces can never produce a motion in any symmetric mode. Thus if we imagine an atom created at the beginning of the world in an antisymmetric mode, it does not matter what has been done to it, or what catastrophes it may have suffered, it will always be antisymmetric. The same is true for any number of electrons or protons. The mutual forces can never create symmetric modes unless there were some

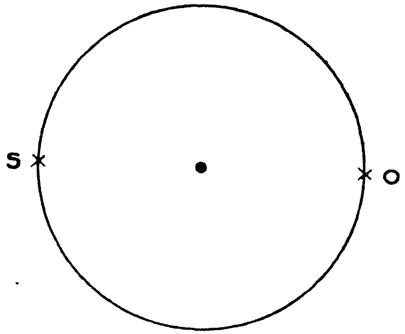


FIG. 46.—SYMMETRY AND ANTISYMMETRY OF ELECTRONS

Two electrons are in the same orbit. There is no antisymmetric mode, and in the symmetric mode, if one is found at *O* the other will most probably be at *S*.

there to start with. Similarly if the world was created with all its atoms only in symmetric modes, then for the rest of time they will continue only in symmetric modes. We can separate up the behaviour of the world into two parts which would go on quite independently of one another.

Now at the beginning of the chapter we saw how the way we think about two electrons is coloured by the idea of their individuality, and we saw that this idea is really unwarranted, because no experiment could ever recognise an individual electron. Thus if we consider the collision of two electrons, instead of saying that one of the tracks may be either electron *A* or else electron *B*, two alternatives, we would like to be able to say that it is simply an electron, with no choice at all. We would like to halve the number of possibilities of the event, and we can now see how this may be done through the complete independence of each other of the symmetric and antisymmetric modes of motion. Our principle then suggests that it is natural to suppose that the world is either wholly symmetric or wholly antisymmetric, but it provides not the faintest hint of which. As far as we may judge, worlds would be perfectly possible of either kind, but here the Exclusion Principle steps in and gives a definite ruling. The world was created antisymmetric.

The process of argument which I have described is roughly the course in which things were discovered, and is an outline of the way the principle is applied even now, but it is very abstruse, and I do not think anyone can really apprehend the idea of antisymmetric motions, except in such simple cases as I have described, without the use of mathematical symbolism. Moreover, there is something very clumsy about the process of first working out events as though the electrons had individuality, and afterwards at a separate stage denying it to them by imposing the condition of antisymmetry. One would like to be able to start with a system which was automatically antisymmetric, and in which from the start there was no question of individuality, but no such process has yet been found. However, it may perhaps help in understanding it if I venture on an account in a rough and ready way, based on the idea that in the antisymmetric motions the electrons tend to avoid one another.

It was one of the earliest of scientific generalisations that two pieces of matter cannot be in the same place at once; the author was, no doubt, some prehistoric philosopher, and he probably reached this momentous conclusion a moment or two before he made the practical verification by being hit on the head with a stone axe. Later philosophers used to regard *impenetrability* as one of the inherent fundamental properties of matter. It was possible to attribute it to repulsive forces between the ultimate particles of matter. But we have to revise our ideas on the subject, because we have seen that the waves of two electrons can certainly overlie one another in the same part of space, and so we must find a new principle to replace the idea of impenetrability. We express it by saying that two electrons may be in the same place, but that they must not be doing the same thing. This statement is intentionally vague, but, roughly speaking, I mean that if an electron is behaving as a wave of some particular kind, then no other electron can be that same kind of wave. For example, if one electron is occupying a normal mode in an atom, then no other may come there. This corresponds to the fact that there is no antisymmetric normal mode for two electrons in the same orbit. Let us now see in a general way how this would work. Electron *A* is carrying on its affairs quite happily alone, when along comes electron *B*. But when *B* gets fairly close, *A* gets uneasy; it fears that *B* is about to commit a trespass, and serves it with a notice by exerting a repulsive force on it. This is the outline of a theory recently put forward by Eddington, which tried to explain the repulsive force between electrons as merely an aspect of the Exclusion Principle. He has not succeeded yet in carrying the theory very far, and it must be confessed that it involves a good many difficulties—for example, we can imagine the existence of a world which is symmetric instead of antisymmetric and yet which has repulsive forces—but still it is an attractive attempt to unify the two aspects of the phenomenon of impenetrability, repulsion, and the Exclusion Principle.

We can still further extend the connection between the Exclusion Principle and the forces of electrons if we take polarisation into account. The rule that forbids electrons from doing the same thing does not in the least prevent them from having identical waves, provided these have opposite polarisations. Let us take into account the magnetic forces which the electron exerts; it is true, as we have seen, that these are not directly observable for free electrons, but they may be permitted in a general view like the present, as being the most natural description we can make of the effects of polarisation. We therefore suppose each electron to have a little bar magnet attached to it. Now the force between two magnets is the result of the attractions and repulsions of their four poles, and these very nearly cancel out. Consequently the magnetic force is very much weaker than the electric at great distances, but as the electrons approach, it increases much more rapidly; whereas the electric force, like gravity, depends on the inverse square of the distance, the force between magnets depends on the inverse fourth power. It may be either attractive or repulsive according to how they are pointing; if they are side by side but pointing opposite ways the force is an attraction. At ordinary distances the magnetic force between two electrons is far less than the electric, but if they are closer together than about 10^{-10} cm. the attractive force of the magnetism will outweigh the repulsion of the electricity.

We will import this fact into our rather fanciful description of how the Exclusion Principle works. When electron *B* is still some way off, *A* begins to get anxious and says, "Keep off, I am doing my business, and I won't have trespassers. I can't quite see what you are doing, but I suspect mischief, and anyhow I can't risk it." He therefore sends out an electric force to drive *B* away. But *B* may be approaching *A* with a rather high speed, so that he gets very close before the electric repulsion can check him. Then *A* says, "Now that I can see you properly, I observe that you are standing on your head, so you certainly won't interfere with me. Come along; this is a nice place," and he exerts all his powers of magnetic attraction to keep *B*. Unfortunately it would violate the conservation of energy for two particles to stick together in free space, so *A* and *B* separate again, after a mutual exchange of momentum. But in the presence of a third body such as a nucleus, this is not so, since there are other means of getting rid of the energy, and it is, in fact, found that in atoms electrons do very generally occur in pairs in the same orbit, but with opposite polarisations. It should be confessed that there is no direct evidence of the magnetic forces between electrons, but it is a well-known fact of chemistry that in most chemical compounds electrons tend to go about in pairs, and our description supplies a hint, if not a reason, why this should be so.

This account of the Exclusion Principle is not to be taken very seriously, since the whole object of the principle is to remove the individuality of the electron, and we have retained it to an even greater extent than in the accepted mathematical theory. It is a dangerous thing to forecast future trends of development, but we may perhaps hope that the theory will ultimately take some such form as the following. As far as concerns the wave aspect there will cease to be electrons at all, but only a sort of electrical juice. It will be the step from wave to particle which automatically cuts the juice into units, and thus it will be the observation that creates the electron. In some such way as this, the electrons would have no individuality from the very start, which is what we would like. If the development should go along these lines, then the current theory may be likened in some ways to the old quantum theory, before the discovery of the wave aspect. That theory was constructed by a blending of pure mechanical principles with the quite foreign idea of quantisation, and it gave a very good account of many things. In the same way we have at present a blending of the mechanical principle of repulsive forces with the quite foreign idea of antisymmetry, and again a number of things can be explained quite well. It must be remembered that these suggestions are entirely speculative, and that no one has yet found any way of advancing beyond the present mixed expression of the idea of impenetrability.

Pauli discovered the Exclusion Principle some time before the advent of the new quantum theory. He derived it from a study of the characteristics of spectra, and expressed it as a rule affecting the various quantum numbers of atoms; for at that time there was no idea of the existence of normal modes or still more of antisymmetry. In this form it explained one of the fundamental questions of chemistry, the Periodic Table of the Elements. It will be convenient in describing this work to adopt the language of the old theory and not to have to speak of antisymmetric modes. In order to be able to do this, we will imagine that at some instant of time the atom is miraculously frozen, so that from being an intricate set of normal modes it becomes a number of little planets describing orbits round a central sun, the nucleus. For example, in the case of the two electrons of Fig. 43 we saw that in fact it would be correct to suppose that they are changing places all the time, but when the atom freezes, one of them continues to describe a little circle and the other a big circle. We must recall that the description of these circles was made by means of three quantum numbers,[1] which defined the nodal surfaces of the vibrations, but that in the particle aspect they could be given a meaning in terms of angular momentum. To the three quantum numbers we must now add a fourth to express the polarisation, for there are two ways of describing any circle, according as the magnet is pointing up or down. The conventional way of describing these two cases is by the numbers ½ and −½. Then every electron in an atom is described by four quantum numbers. Now we saw that in the special case where two electrons were in the same orbit, there is no antisymmetric mode, but only a symmetric one, and this last is barred by the Exclusion Principle. It follows that in the present method of description no two electrons in the atom may have all their four quantum numbers the same. This provides the key to the structure of atoms, and we will proceed to consider how the Table of the Elements is built up on these lines.

The first element is hydrogen with atomic number 1.

---

[1] See pp. 116–121.

It can therefore only hold one electron. We have seen that there are an indefinite number of possible orbits for this electron, but the atom will tend to radiate away energy, and so the electron will sink to the lowest energy level that exists. This is the level 0, 0, 0, ½, or the fourth number may alternatively be −½ without affecting the energy. Such an atom will behave like a magnet, and this is verified by the fact that in the experiment of Stern and Gerlach a stream of hydrogen atoms is split into two, just like silver. Atomic hydrogen is not a very familiar substance, since the atoms tend to combine in pairs into molecules; but the atomic form can be manufactured, and will continue to exist for the fraction of a second, which is long enough to make the study of its properties possible.

The next element is helium with atomic number 2. In their state of lowest energy the two electrons get into the levels 0, 0, 0, ½, and 0, 0, 0, −½, which they may do since the fourth quantum numbers are different. In such an atom the magnetic forces of the two electrons cancel out, so that in the Stern-Gerlach experiment a beam of helium atoms is undeflected. The helium atom is the most symmetrical thing in the world; all directions in it are completely equivalent to one another.

Lithium has atomic number 3, and so now we have to fit in a third electron. The first two occupy the places 0, 0, 0, ½ and 0, 0, 0, −½, and the Exclusion Principle demands that the third should have numbers at least one of which is different from these sets. As the fourth number is only allowed the above two values, one of the others must be changed. Consequently the third electron takes numbers 1, 0, 0, ½. In the model where the electron is represented by a particle describing an orbit, this would be a rather eccentric ellipse. Lithium, like hydrogen, shows the Stern-Gerlach effect.

For beryllium, with atomic number 4, the next electron takes the orbit 1, 0, 0, $-\frac{1}{2}$, but the next element, boron, can only accommodate its fifth electron by altering the second number. This electron will therefore be 0, 1, 0, $\frac{1}{2}$. In this sort of way the addition of each new electron occupies another set of quantum numbers, and as they are filled, the elements take on a succession of different chemical characters. For instance, the gas neon has 10 electrons, and we may give a list of its electrons by recalling the rules given on p. 116. They will be

$$\begin{gathered} 000\tfrac{1}{2},\quad 100\tfrac{1}{2},\quad 010\tfrac{1}{2},\quad 011\tfrac{1}{2},\ 01{-1}\tfrac{1}{2} \\ 000{-\tfrac{1}{2}},\ 100{-\tfrac{1}{2}},\ 010{-\tfrac{1}{2}},\ 011{-\tfrac{1}{2}},\ 01{-1}{-\tfrac{1}{2}}. \end{gathered}$$

There is now no more room at such a low energy level, and so the eleventh electron of the next element, sodium, is forced to take the numbers 2, 0, 0, $\frac{1}{2}$, representing an eccentric ellipse. Such an atom, with a completely balanced core and a single outside electron, has obvious resemblances to lithium of which the same was true, and so it is not surprising that in their chemical behaviour sodium and lithium are very similar.

We will now turn aside to other applications of the Exclusion Principle. We have seen how the chemistry of the elements proves the principle true for electrons, and it is natural to ask whether it is true for protons also. The evidence here is not so easy to come by, but it is definitely established that protons have the same antisymmetric character as electrons. In Chapter VI we described how hydrogen gas must be regarded as a mixture of two separate gases, and we gave a rough description of their difference as depending on the directions of polarisation of the two protons. The proper description depends on the condition that the two protons must be in an antisymmetric mode.

Another application of the principle is connected with collisions. When a moving electron *B* collides with a fixed electron *A*, allowance must be made for their loss of individuality in the collision. The scattered *A* of the left-hand diagram of Fig. 40 is capable of undergoing a peculiar kind of *interference* with the *B* of the right-hand diagram, even though in the particle aspect the two diagrams belong to different types of collision. This interference in no way affects the rules concerning the directions and velocities of the two electrons in a single collision, since these are governed by the conservation theorems; but, as was shown by Mott, there is a very pronounced effect on the probability of any pair of directions occurring. Consider the collisions in which the electrons both go off at 45° to the direction of *B*'s incident motion; these will be only *half* as frequent as they would be if the condition of antisymmetry were disregarded. There is no very good experimental evidence on the point, but what there is supports the theoretical prediction. The theory is much more thoroughly verified by the study of the collisions of α-particles with helium nuclei (which of course are also α-particles). The rule for a pair of α-particles is quite different from that for a pair of electrons. An α-particle is composed of four protons and two electrons, and, in consequence of the even numbers of both protons and electrons, it can be shown that the rule for the two particles, regarded as wholes, is that they must be in a *symmetric* mode. This makes a great difference in the probability of the various types of collision. Thus the 45° collisions are now *twice* as frequent as they would be if there had been no Exclusion Principle, and at certain other angles the collisions will be rarer. These results have been verified experimentally in a most satisfactory manner.

Turning to other questions, one of the triumphs of the Exclusion Principle has been in connection with the theory of magnetism. This was due to Heisenberg. For a long time it had been perceived that there were serious difficulties in connection with the very strong magnetism of iron. It was known that the magnetism must be due to an enormous number of little magnets all pointing the same way, and it was known that these magnets were in fact electrons, but there was no way of explaining why the electrons should all want to point the same way. The mutual forces of the magnets could be worked out, and totally failed to provide as much as a thousandth part of the forces necessary to make a sufficient number of electrons point in the same direction. But now let us see how the Exclusion Principle can help. Consider the interaction of three electrons in the rather fanciful way I have used above. *A* and *B* are behaving in certain ways, carefully avoiding one another, and along comes *C*. As it approaches, we expect that *A* and *B* will each serve it with a notice of trespass by repelling it. But *C*, of course, is also telling *A* and *B* that they are in danger of trespassing, and they will be set in motion in consequence. In being pushed away, *A* has not only to avoid *C*, but it has also to avoid *B*, which is being pushed away from *C* too. The ordinary forces of repulsion prevent the trespass between a pair of electrons, but they are not adequate to prevent the trespass of *A* and *B* when *C* is also interfering. This is, of course, not a proof, but it does suggest how it comes about that the joint effects of *A*, *B*, and *C* are not simply composed of the mutual effects between each of the pairs *A* and *B*, *B* and *C*, *C* and *A*. Heisenberg attacked the question with all the proper machinery of antisymmetric waves, and showed that in fact the force acting on each of the electron-magnets was not simply the sum of the forces from all the rest, and that under certain conditions it could be enormously greater. This is sufficient to explain the magnetism of iron, but the theory is still rather incomplete, because no one knows why it is only iron and a few other substances that show this strong magnetism. In a general way, however, Heisenberg's work draws attention to the enormously important fact that the joint influences between a number of bodies cannot be estimated by simply taking them together in pairs.

Another application of the Exclusion Principle is to the theory of metals, and this was made by Sommerfeld. Before this there had been very severe difficulties in understanding how a metal conducts electricity. The whole subject of electric conduction has been studied experimentally from many different points of view, and quite a dozen different properties of it are known, so that the difficulty was not so much in explaining a single phenomenon as in choosing an explanation for one which would not thereby give wrong results for all the others. I cannot describe all the different phenomena, but must select one cardinal difficulty which arose in most of the older attempts at explanation. The obvious way of explaining electric conduction is to suppose that in a metal each atom sets free an electron or two, and that these electrons move freely between the planes of the metal crystals; under an electric force they will drift down the wire, and their motion represents the current. Some such explanation of electric conduction is almost inevitable, but on the older theories it gave insuperable difficulties in connection with the specific heat of the metal.

To understand the way in which these difficulties arise, we must consider a branch of physics on which I have hardly touched. The temperature of any body is determined by the energy of its constituent atoms, and the average energy of each of them is the same, whether they are light or heavy. Every constituent particle that is capable of moving claims its share of the energy. If the substance is a metal, not only should every atom take its share, but also every free electron. Now the experimental measure of specific heat shows that when a copper wire is heated from say the freezing-point of water to the boiling-point, the energy absorbed is just enough for the atoms, and none is required for the free electrons. Thus the electric conduction tells us the electrons are free and the specific heat that they are not free, and we have an insuperable contradiction. This was removed by the Exclusion Principle. If we suppose the electrons at first all bound to the atoms, and if we then free them one by one, we get a state of affairs much like what we had in constructing a single atom. The freed electrons are to be regarded as waves pervading the whole space of the metal crystal. The first few take the form of waves of low energy, but the later ones find these modes of low energy occupied, and so have to take higher places. The whole metal is definitely *filled* with electrons, so that no one of them can change its mode without trespassing on the rest. Moreover, some of these electrons must have quite large energy, because there is no room below, and it turns out that this energy is much greater than that of the atoms. Consequently in the slight heating of the metal from the freezing to the boiling-point of water there is no demand for energy for the electrons, since they already have far more than the atoms. It has been calculated that this is true for temperatures up to about ten thousand degrees; above that the energy of the atoms would become so great that the electrons could acquire further energy from them, and consequently the specific heat would increase. Since, however, all metals have dissolved into vapour at much lower temperatures, this does not seem a very important exception.

To conclude this branch of the subject, I may mention that the same phenomenon can happen for atoms, only at an enormously higher density. It was shown by Fowler that in the interior of those peculiar stars that are known as white dwarfs, the atoms are behaving in much the same way as electrons ordinarily do in a metal.

In the preceding chapters I have attempted to give a general description of how things behave, and have put the emphasis in places which does not accord very well with the historical course of development. If the literature of the last twelve years is consulted, it will be found that a quite preponderating majority of the papers has been concerned with the structure of atoms and with spectra. These subjects involve many fascinating problems, but lead to such intricate detail that it would hardly have been possible to describe them here. I did not want to describe them, for I think that the physicist can now free his edifice from the scaffolding which was necessary in its construction. The scaffolding remains, but as a new branch of science, and, freed from it, the science of physics has returned to its prime purpose, the study of the ultimate constituents of the world. I hope I have shown that it has had great successes in this study. The new views are curiously different from the crude views of human intuition, but I hope that I have not totally failed in showing that they are sensible views; and that, with all its difficulty, there is nothing foreign to our processes of thought in the New Conceptions of Matter.

INDEX

ÆTHER, 22.
α-particles, 14, 157, 186.
amplitude, 31
angular momentum, 26, 84, 119.
antisymmetry, 171–182.
Aston, 17.
atomic number, 15.

BEATS, 169.
β-particles, 14.
Bohr, 80, 118, 142, 144.
de Broglie, 62, 71, 107, 121

CAUSALITY, 101.
circular polarisation, 128.
Compton, 95, 160.
crystals, 68, 126.

DALTON, 9
Davisson, 72, 85.
determinism, 101
diffraction, 44.
Dirac, 136, 144.
double refraction, 125–131.

EDDINGTON, 3, 180.
Einstein, 23, 81, 163.
electric conduction, 188.
electrolysis, 10.
electron, 13, etc.
elements, 9, 183.
energy, 27, 84, 89, etc.
energy level, 122, 132.
exchange, 170–175.

FARADAY, 10.
Fowler, R. H., 190.
frequency, 31, 84.
Fresnel, 22, 125.

GAMOW, 157.
Gerlach, 136-141.
Germer, 72.
Goudsmit, 134.
grating, 63–67, 104.
group-velocity, 49.

HALF-WORLDS, 83, 155.
Hamilton, 55–62, 107.
harmonic waves, 30–35.
Heisenberg, 80, 124, 159, 187.
Huygens, 38, 43, 59, 125.
hydrogen atom, 108, 184.
hydrogen gas, 144, 185.
hyper-fine structure, 145.

INTENSITY, 35, 100.
interference, 39.
isotopes, 17.

v. LAUE, 68.
least action, 54–58.
Lebedev, 95.
light pressure, 25, 93.
Lyman, 123.

MAGNET, 137, 180.
magnetism, 187.
Maupertuis, 55.
microscope, 91, 96.
momentum, 25, 84, etc.
Mott, 143, 157, 186.

NEWTON, 24, 53, 125.
nodes, 37, 84, 111–115.
normal modes, 110.
nucleus, 15.

PAULI, 159, 182.
phase, 32.
photo-electricity, 163.
photon, 20, 130.
Planck, 84.
probability, 12, 100.
proton, 18, 144.

QUANTUM, 62, 84, 119.

RAMAN, 167.
Ramsauer, 158.
Rayleigh, 166.
refraction, 59, 166.
resonance radiation, 165.
Röntgen, 67.
Rutherford, 14, 19.
Rydberg, 117.

SCALE MODEL, 11.
Schrödinger, 109.
scintillation, 79.
Sommerfeld, 188.
specific heat, 189.
spinning electron, 134, 142.
standing waves, 37.
Stern, 136–141.
superposition, 36.

THOMSON, G. P., 72, 84.
Thomson, J. J., 13, 162.
transmutation, 19.

UHLENBECK, 134.

WAVE-LENGTH, 31, 84, etc.
wave-packet, 78, 87.
wave-velocity, 31.
white light, 105.

X-RAYS, 67–74.

YOUNG, 22, 40, 42.

ZEEMAN, 133.